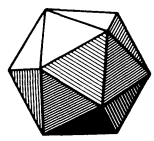
THE AMERICAN MATHEMATICAL
MONTHLY



Volume 106, Number 1 January 1999

AN OFFICIAL PUBLICATION OF THE MATHEMATICAL ASSOCIATION OF AMERICA

NOTICE TO AUTHORS

The MONTHLY publishes articles, as well as notes and other features, about mathematics and the profession. Its readers span a broad spectrum of mathematical interests, and include professional mathematicians as well as students of mathematics at all collegiate levels. Authors are invited to submit articles and notes that bring interesting mathematical ideas to a wide audience of MONTHLY readers.

The MONTHLY's readers expect a high standard of exposition; they expect articles to inform, stimulate, challenge, enlighten, and even entertain. MONTHLY articles are meant to be read, enjoyed, and discussed, rather than just archived. Articles may be expositions of old or new results, historical or biographical essays, speculations or definitive treatments, broad developments, or explorations of a single application. Novelty and generality are far less important than clarity of exposition and broad appeal. Appropriate figures, diagrams, and photographs are encouraged.

Notes are short, sharply focussed, and possibly informal. They are often gems that provide a new proof of an old theorem, a novel presentation of a familiar theme, or a lively discussion of a single issue.

Articles and Notes should be sent to the Editor:

ROGER A. HORN
1515 Mineral Square, Room 142
University of Utah
Salt Lake City, UT 84112

Please send your email address and 3 copies of the complete manuscript (including all figures with captions and lettering), typewritten on only one side of the paper. In addition, send one original copy of all figures without lettering, drawn carefully in black ink on separate sheets of paper.

Letters to the Editor on any topic are invited; please send to the MONTHLY's Utah office. Comments, criticisms, and suggestions for making the MONTHLY more lively, entertaining, and informative are welcome.

See the MONTHLY section of MAA Online for current information such as contents of issues, descriptive summaries of forthcoming articles, tips for authors, and preparation of manuscripts in TEX:

http://www.maa.org/

Proposed problems or solutions should be sent to:

DANIEL ULLMAN, MONTHLY Problems
Department of Mathematics
The George Washington University
2201 G Street, NW, Room 428A
Washington, DC 20052

Please send 2 copies of all problems/solutions material, typewritten on only one side of the paper.

EDITOR: ROGER A. HORN
monthly@math.utah.edu

ASSOCIATE EDITORS:

WILLIAM ADKINS	VICTOR KATZ
DONNA BEERS	STEVEN KRANTZ
HAROLD BOAS	JIMMIE LAWSON
RICHARD BUMBY	CATHERINE COLE McGEOCH
JAMES CASE	RICHARD NOWAKOWSKI
JANE DAY	ARNOLD OSTEBEE
JOHN DUNCAN	KAREN PARSHALL
PETER DUREN	EDWARD SCHEINERMAN
GERALD EDGAR	ABE SHENITZER
JOHN EWING	WALTER STROMQUIST
JOSEPH GALLIAN	ALAN TUCKER
ROBERT GREENE	DANIEL ULLMAN
RICHARD GUY	DANIEL VELLEMAN
PAUL HALMOS	ANN WATKINS
GUERSHON HAREL	DOUGLAS WEST
DAVID HOAGLIN	HERBERT WILF

EDITORIAL ASSISTANTS:

ARLEE CRAPO
MEGAN TONKOVICH

Reprint permission:
DONALD ALBERS, Director of Publication

Advertising Correspondence:
Mr. JOE WATSON, Advertising Manager

Change of address, missing issues inquiries, and other subscription correspondence:
MAA Service Center
maahq@maa.org

All at the address:

The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC 20036

Microfilm Editions: University Microfilms International, Serial Bid coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT. Copyrighted by the Mathematical Association of America (Incorporated), 1999, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. "Permission to make copies of individual articles, in paper or electronic form, including posting on personal and class web pages, for educational and scientific use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear the following copyright notice: [Copyright the Mathematical Association of America 1999. All rights reserved.] Abstracting, with credit is permitted. To copy otherwise or to republish, requires specific permission of the MAA's Director of Publication and possibly a fee." Second class postage paid at Washington, DC, and additional mailing offices. ***Postmaster***: Send address changes to the American Mathematical Monthly, Membership / Subscription Department, MAA, 1529 Eighteenth Street, N.W., Washington, DC, 20036-1385.

Large Torsional Oscillations in Suspension Bridges Revisited: Fixing an Old Approximation

P. J. McKenna

1. INTRODUCTION. In *Inventions and Technology*, Frederic D. Schwartz writes "The Tacoma Narrows Bridge collapse is technology's version of the JFK assassination. There's the grainy black-and-white film endlessly scrutinized frame by frame; the reams of expert analysis next to impossible for a layperson to evaluate; and, of course, the buffs who are convinced that only they know the real story"[29].

He was referring to a "genteel professorial catfight" [28] between some engineering writers and me about the explanation of the famous destructive large amplitude torsional oscillation captured on film on November 7th, 1940, and now a staple in physics classes.

Until the collapse of the Tacoma Narrows bridge, it seems that most suspension bridge building had been largely by rule of thumb, without a theoretical underpinning. As a result, early bridges, including the Golden Gate bridge and the Tacoma Narrows bridge, exhibited interesting behaviour:

1. they were prone to large-scale oscillation, both torsional and purely vertical
2. occurrence of this large-scale oscillation is dependent on initial conditions and can be started by a single gust [25]
3. large vertical oscillations could rapidly change, virtually instantaneously, to torsional oscillation
4. these bridges exhibit localised travelling wave solutions
5. for small oscillation, the observed behaviour is almost perfectly linear.

Good historical sources for these behaviours are [2] and [5].

Of course, these behaviours strongly suggest nonlinearity, and in the late eighties, following recent and rapid progress in nonlinear analysis of boundary value problems that had just been taking place, I and co-workers attempted to give a mathematical explanation for these phenomena [13], [16–19]. Our starting point was that an unloaded cable cannot be described by the usual Hooke's law, since it resists expansion but not compression. Thus if an unloaded cable is expanded downward by a distance u from the unloaded state, the cable should have a resisting force ku^+, in other words, ku if u is positive, and 0 if u is negative. When the cable is loaded, it stretches to a new equilibrium, around which it obeys the linear Hooke's law until deviations from equilibrium become large enough to result in loss of tension when the cable approaches the unloaded state.

Using the tools of nonlinear analysis and numerical investigations, we showed that an equation for a nonlinearly supported beam with this type of nonlinearity could explain both large amplitude vertical periodic oscillations and the travelling wave behaviour. With the usual fervor of the newly converted, I suggested that the

nonlinearity induced by the alternate slackening and tightening of the cables could also explain the more famous large amplitude torsional oscillations.

The purpose of this paper is to explore a startling and different possibility. If one is interested solely in a simple version of the Tacoma Narrows oscillation, namely large amplitude torsional oscillation about equilibrium, then the puzzle may have its roots in a simple trigonometry approximation introduced in the engineering literature fifty years ago [**5**], and unwittingly reproduced ever since, for example in [**1**].

After the collapse, the initial response of the engineering community, which has served well over the years, was to develop a theory of small oscillations and to construct bridges so that the oscillations stayed in that range. This theory involved many linearising approximations.

The theory was successful in the practical sense. Newer bridges no longer engaged in interesting large amplitude behaviour. And older ones like the Golden Gate were soon modified so that they didn't either.

Of course, when one makes the small angle and other near-equilibrium assumptions, a theory emerges that cannot explain the large-amplitude motions. Redesigning the bridge to remove the offending behaviour is not the same as mathematically understanding its cause.

This is what the distinguished civil engineering writer, the late Mario Salvadori, meant when he wrote to me "... having found obvious and effective physical ways of avoiding the problem, engineers will not give too much attention to the mathematical solution of this fascinating puzzle and [I] am delighted to learn that mathematicians like you are interested in it" [**22**].

In this paper, we re-derive a mechanical model for a beam or plate oscillating torsionally about equilibrium, and suspended at both sides or ends by cables. We show how the 'small-angle' linearisations can remove a large class of large-amplitude nonlinear solutions that can be sustained by extremely small periodic forcing terms.

Based on the original report of the Tacoma Narrows disaster [**2**], we choose appropriate values for the physical constants in the system of differential equations.

Then we explore the implications and consequences of making, or not making, the trigonometry approximation, and show that even if no loss of tension in the cables is assumed, the nonlinearities introduced by the geometry of the situation are sufficient to explain the large amplitude oscillations seen before the collapse. The large amplitude oscillations that can result from small forcing terms exist over the right range of frequency and amplitude to match the historical observations.

The last nonlinear effect that we listed and would like to explain is the virtually instantaneous change from vertical to torsional oscillation. In Section 4, we explore some of the consequences of assuming that the cables briefly lose tension in a transient way. We observe numerically that when purely vertical oscillations are large enough to allow the supporting cables to lose tension, and the system is subjected to tiny torsional perturbations, the system becomes violently unstable in the torsional dimension.

2. SETTING UP THE MODEL. There is nothing controversial in this section. We set up the equations for vertical and torsional motion of a rod, (or plate), suspended by springs at both ends, (or sides), and free to move vertically and rotate about its center of gravity. In the first subsection, we derive the equations

from the geometry of the situation, and in the second, we choose reasonable values for the physical constants and the forcing terms.

2.1 The equations for vertical and torsional oscillations. If a spring with spring constant K is extended by a distance y, the potential energy is $Ky^2/2$. If a rod of mass m and length $2l$ rotates about its center of gravity with angular velocity $\dot{\theta}$, then, its kinetic energy is given by $(1/6)ml^2(\dot{\theta})^2$ [**31**, p. 202]. Assume the rod is suspended as in Figure 1 by springs that resist expansion with a spring constant K at each end. Let y denote the downward distance of the center of gravity of the rod from the unloaded state. Let θ denote the angle of the rod from the horizontal. Let y^+ be the positive part of y, that is, $y^+ = \max\{y, 0\}$.

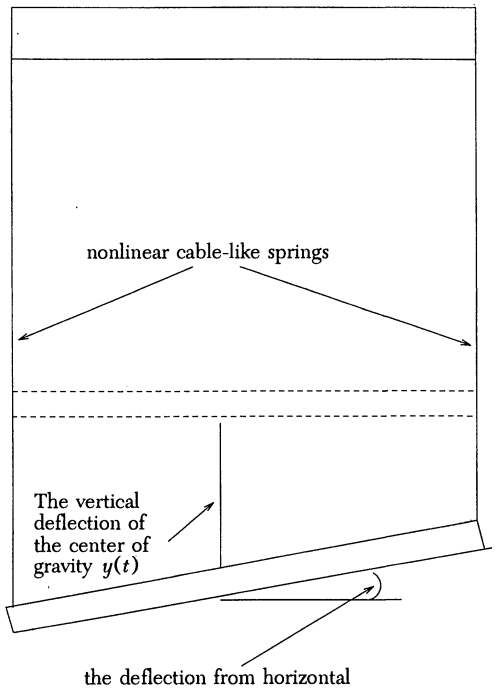


Figure 1. A cross-section of the bridge as an inflexible rod supported by two springs at each side, giving rise to the system of equations (3) and (4) and simplifying to (5) and (6) when there is no loss of tension.

The potential energy due to gravity is $-mgy$. The extension is $(y - l\sin\theta)^+$ in one spring and $(y + l\sin\theta)^+$ in the other. Actually, this involves another hidden small-angle assumption, namely that the cables remain vertical, although the lateral motion of the ends of the plate or rod deflects them slightly to the left or right. Thus, this model actually overstates slightly the resistance to torsional motion. When the cables are of length 100 feet and the lateral motion is about six feet, as they were in the Tacoma Narrows bridge, this seems a reasonable approximation.

Thus, the total potential energy is

$$V = (K/2)\Big(\big((y - l\sin\theta)^+\big)^2 + \big(y + l\sin\theta)^+\big)^2 - mgy$$

and the total kinetic energy is

$$T = m\dot{y}^2/2 + (1/6)ml^2\dot{\theta}^2.$$

Let $L = T - V$ and put $(d/dt)(\delta L/\delta\dot{\theta}) = (\delta L/\delta\theta)$ and $(d/dt)(\delta L/\delta\dot{y}) = \delta L/\delta y$. We obtain the equations

$$(1/3)ml^2\ddot{\theta} = (Kl)\cos\theta\big((y - l\sin\theta)^+ - (y + l\sin\theta)^+\big) \tag{1}$$

$$\ddot{y} = -K\big((y - l\sin\theta)^+ + (y + l\sin\theta)^+\big) + mg \tag{2}$$

Since the springs are assumed to remain vertical, the force exerted by the spring is not in the torsional direction perpendicular to the rod but is at an angle θ to that perpendicular. This is why the additional term $\cos\theta$ occurs in the torsional equation from the springs.

Simplifying and adding a small viscous damping term $\delta\dot{\theta}$ to the first equation and $\delta\dot{y}$ to the second, and adding an external forcing term $f(t)$ (to be determined later), to the torsional equation, we end up with the system

$$\ddot{\theta} = -\delta\dot{\theta} + (3K/ml)\cos\theta\big[(y - l\sin\theta)^+ - (y + l\sin\theta)^+\big] + f(t) \tag{3}$$

$$\ddot{y} = -\delta\dot{y} - (K/m)\big[(y - l\sin\theta)^+ + (y + l\sin\theta)^+\big] + g \tag{4}$$

Finally, if we assume the cables never lose tension so that $(y - l\sin\theta)^+ = (y - l\sin\theta)$ and $(y + l\sin\theta)^+ = (y + l\sin\theta)$, we end up with the uncoupled equations

$$\ddot{\theta} = -\delta\dot{\theta} - (6K/m)\cos\theta\sin\theta + f(t) \tag{5}$$

for the torsional motion and

$$\ddot{y} = -\delta\dot{y} - (2K/m)y + g \tag{6}$$

for the vertical motion.

Now, we are in a position to introduce the error that we believe is key to the failure to understand the large amplitude torsional oscillation in the case of the Tacoma Narrows bridge. If we are interested in studying *very small* oscillations, we could call it an approximation, but for large oscillations, it is an error.

We put $\sin\theta = \theta$ and $\cos\theta = 1$. This gets us to the two equations

$$\ddot{\theta} = -\delta\dot{\theta} - (6K/m)\theta + f(t) \tag{7}$$

and

$$\ddot{y} = -\delta\dot{y} - (2K/m)y + g \tag{8}$$

This is the point at which the discussion of torsional oscillation starts in the engineering literature. For example, in [**26**] one finds torsional motion described by

$$I\big(\ddot{\alpha} + \delta\dot{\alpha} + \omega_\alpha^2\alpha\big) = f\dot{\alpha}, \tag{9}$$

where α is the torsional angle and $f\dot{\alpha}$ is an external force of aerodynamic origin. In [**4**], this equation is used to describe the motion of the Tacoma Narrows bridge prior to the collapse. Even in recent engineering literature [**1**], which claims to derive "the coupled equations of motion in their most general and nonlinear form", this same mistake is reproduced, although "... one has to expect responses of large vibrational amplitudes which necessitates a nonlinear formulation as presented in this paper."

In the pioneering studies of small oscillations of suspension bridges, summarised in [**5**], it is clear why this approximation was made. There was simply no

technology available to solve nonlinear equations like (5). There was no choice but to linearize the equation and hope the errors introduced were not too big. Indeed, experience showed that for *very small* oscillations, this was reasonable.

However, today we can solve (5) accurately. Thus we can determine whether there is an important difference between the trigonometrically correct model and the linear one. But before we do this, we should choose appropriate constants. This is the task of the next subsection.

2.2 Choosing physical constants and external forcing term. Our main source is [2]. Since the model is very simple, we do not concern ourselves with very precise choice of constants but content ourselves with getting the magnitudes about right. The mass of a foot of the bridge was about 5,000 lbs. so we choose the m in our equations to be 2,500 kgs. The width of the bridge was about 12 meters so we choose l to be 6.

The bridge would deflect about .5 meters when loaded with 100 kgs. for each .3 m. of bridge. Since there are two springs, this gives the equation $2K(.5) = 100(9.8)$ for $2Ky = mg$, and we choose $K = 1000$.

For the external forcing term, we choose $f(t) = \lambda \sin \mu t$ and investigate the response of the different equations to this type of periodic forcing. The frequency of the motion of the bridge before the collapse was about 12 to 14 cycles per minute so we take μ between 1.2 to 1.6.

There does not seem to be much information about the *amplitude* of the forcing term, although there was an attempt to measure it in [**26**]. Apparently, one can measure small oscillations caused by forces that induce oscillations of about 3 deg. "It was found necessary to permit only small amplitudes of oscillation to occur, (e.g. in torsion $0 \leq \alpha \leq \pm 3$ deg)" [sic]. Thus, we choose λ small enough to create oscillation of this order of magnitude in the linear model.

Finally, there is a consensus that the term describing the viscous damping should have a coefficient of about .01 [**2**].

We now have a system that, in the absence of forcing, settles down to an equilibrium with $y = 12.25$ and $\theta = 0$. As long as the oscillation is such that the $y \pm l \sin\theta$ remains below zero, i.e., the deflections upward from equilibrium do not exceed about 12 meters, the equations remain uncoupled, and we investigate the long-term behaviour of the forced pendulum equation (5) versus the linearized version (7) with these constants and a variety of initial conditions and small forcing terms.

2.3 What happened at Tacoma Narrows. We recall what happened before the collapse. Again, our source is [**2**].

The bridge engaged in vertical oscillations even during construction. "*Prior to 10.00 A.M. on the day of the failure, there were no recorded instances of the oscillation being otherwise than with the two cables in phase and with no torsional motion.*"

On November 7, for some time before 10:00 AM, the bridge was engaged in what seemed like normal vertical motions, with amplitudes of about 5 feet and a frequency of about 38 per minute. The motion was apparently more violent than usual, with eight nodes. Then the torsional motion began, as the bridge was being observed by Professor F. B. Farquharson [**2**]:

> ... a violent change in the motion was noted. This change appeared to take place without any intermediate stages and with such extreme violence that the span appeared to be about to roll completely over.

> The motion, which a moment before had involved a number of waves, (nine or ten) [This means a eight-or-nine noded vertical motion], had shifted *almost instantly* [our emphasis] to two [a one-noded torsional motion].
>
> At the moment of first observing the main span, . . . , the motion had a frequency of 14 cycles per minute. . . . the node was at the center of the main span and the structure was subjected to a violent torsional action about this point.
>
> At times, for a short period, the motion changed over to a single wave on each cable but still with the cables out of step. This motion, which never lasted long seemed to be of slightly greater amplitude than the single-noded motion, but of the same frequency. [This motion, of double amplitude of 28 feet continued for approximately forty-five minutes.]

There is no consensus on what caused the sudden change to torsional motion. In [**23**, p. 209], this transition is explained as "some fortuitous condition broke the bridge action."

It may have been a minor structural failure, enough to jar the bridge in a torsional direction. We explore the consequences of just such a single push on the two models we compare: the linear and the trigonometric ones. Thus, it is our task to look for large-amplitude periodic motions that are primarily torsional with an amplitude of about ± 1 radian, corresponding to a small torsional forcing term with a frequency in the neighbourhood of 12 to 14 cycles per minute.

2.4 A few words about forcing. We have to choose some method to model the aerodynamic and other forces acting on the bridge that cause it to go into a periodic motion. There is really no way to say precisely what the forces would be when a huge structure is oscillating up and down by 28 feet every five seconds.

According to [**26**], for a cross-section similar to the Tacoma Narrows bridge in a wind tunnel, the aerodynamic forces induced approximately sinusoidal oscillations of amplitude of plus or minus three degrees. In all our experiments, we explore the response of the system to a sinusoidal forcing term of the form $\lambda \sin(\mu t)$. We then choose the value of λ in order to induce oscillations of three degrees near equilibrium in the linear model.

In this sense, we have chosen the particular form $\lambda \sin(\mu t)$ as a 'generic' oscillatory force, with the right frequency and amplitude. The conclusions would be the same for any other oscillatory force of roughly the same magnitude and frequency. For example, we repeated some of these experiments for the forcing term $\lambda(\sin(\mu t))^3$. The results were qualitatively the same: small forcing could give rise to either large or small periodic long-term behaviour, with the eventual outcome dependent on the initial conditions.

3. NUMERICAL EXPERIMENTS: THE EFFECT OF NOT LINEARIZING. We now investigate the response of the two equations to various initial conditions and forcing terms. Our first equation, with the correct trigonometry is

$$\ddot{\theta} = -0.01\theta - (2.4)\cos\theta \sin\theta + \lambda \sin \mu t, \tag{10}$$

which when linearized becomes

$$\ddot{\theta} = -0.01\dot{\theta} - (2.4)\theta + \lambda \sin \mu t. \tag{11}$$

3.1 Long-term behaviour with and without The Error. We start with initial conditions that mimic a large torsional push. We choose $\theta(0) = 1.2$ and $\dot{\theta}(0) = 0$, and start with $\mu = 1.2$ and $\lambda = 0.06$. We run the initial value problem for one thousand periods to see what the system has settled down to. This is shown in Figure 2.

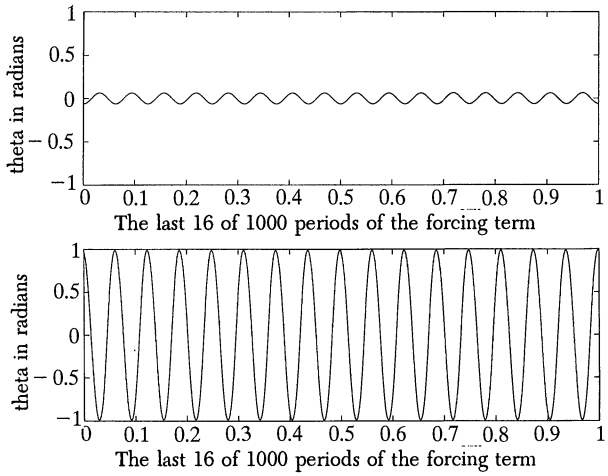


Figure 2. The difference in long-term outcomes depending on whether we solve the trigonometrically correct equation (10) or the linearized equation (11) with a large torsional push and a small periodic forcing term.

On the top of Figure 2, the linear oscillator has settled to an oscillation of approximately ± 3 deg. Since this is true of all experiments we run, we do not repeat the picture, but simply remark that in the linear oscillator, the oscillation has died off.

The bottom half of Figure 2 shows that the torsional oscillator of equation (10) has settled into a periodic oscillation of *large amplitude*. The single large push at the start of the experiment has induced a permanent large amplitude torsional oscillation. This solution represents a torsional oscillation of about one radian, with a period of about 5.2, and a vertical amplitude at the sides of about the correct amount of 10 meters. The period is a little larger than the range of 4.25–5.00 reported in the bridge, but it is certainly a reasonable approximation.

Now we solve the initial value problem for the trigonometric oscillator with the initial conditions $\theta(0) = 0$ and $\dot{\theta}(0) = 0$. This time, it eventually settles down to the small periodic oscillation shown in Figure 2. The results of solving the trigonometric oscillator for large and small torsional pushes is contrasted in Figure 3. If the initial values remain small, one can end up in what is basically the linear situation: compare the top graphs in Figures 2 and 3. On the other hand, the large torsional push can result in the large amplitude oscillation shown in the bottom graphs of Figures 2 and 3.

This is the phenomenon that we wish to emphasize: the trigonometric oscillator can have *several different periodic responses to the same periodic forcing term. Which*

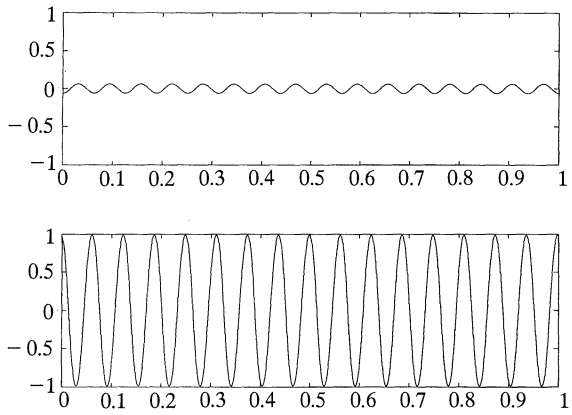


Figure 3. Eventual behaviour of the trigonometrically correct oscillator combining a large torsional push (bottom) or small (top), both with a small torsional forcing term with μ at 1.2.

response eventually results can be determined by a simple transient event such as a large single push.

Let's explore this a little more. We take $\mu = 1.3$. Also, take $\lambda = 0.02$, enough to induce an oscillation in the linear oscillator of ± 1 deg. Needless to say, even after the large push, the linear oscillator settles down after a large time to near-equilibrium, as seen in Figure 4. Here, the period, matching that of the forcing term, is a more realistic 4.83. Again, the large push induces a permanent large torsional oscillation in the trigonometric oscillator, of slightly smaller amplitude than the earlier case, but still huge.

The same results occur if we take $\mu = 1.4$, where we get a large amplitude oscillation with a period of 4.5, corresponding closely to that reported (about 4.3) at the start of the torsional oscillation on the bridge. Later, it seems to have slowed down to 5.0.

Similar results were obtained for $\mu = 1.5$ but they disappear above this value until one approaches values of μ that are integer multiples of the values we have been discussing. At these values subharmonic behaviour occurs, as we discuss in Section 3.2.

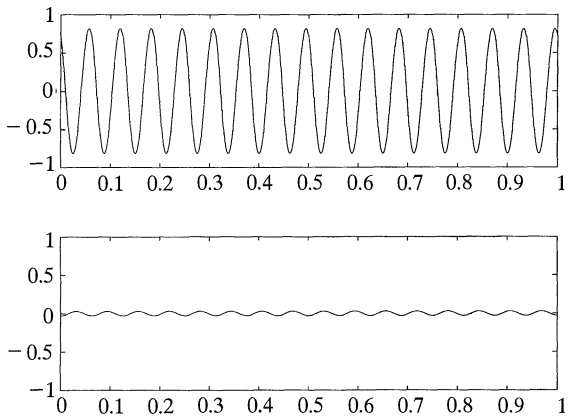


Figure 4. Multiple periodic solutions similar to Figure 3, but with μ changed to 1.3.

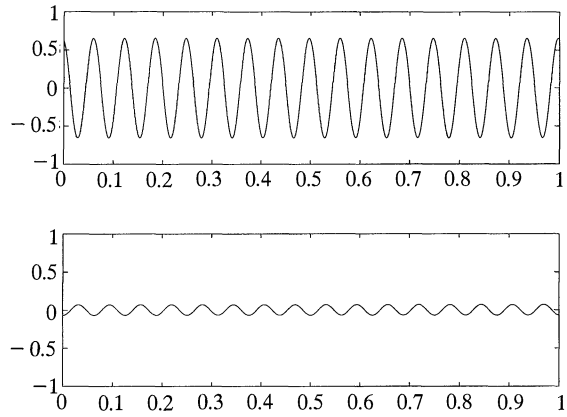


Figure 5. Multiple periodic solutions similar to Figure 3, but with μ changed to 1.4.

This is the most important conclusion of this article: over a range of frequency that is close to that observed before the collapse of the Tacoma Narrows, the final oscillation that results in the trigonometric oscillator after a large time can be either small or very large. All it may take is a single push to change the eventual outcome from small near-equilibrium behaviour to large torsional periodic motions.

3.2 Subharmonic responses. Anyone familiar with nonlinear dynamics who has read this far doubtless wonders why we are emphasizing responses of the same period as the forcing term. In fact, we expect that the periodic solutions of (10) discussed in the previous section can be sustained, not just by the forcing term of the same period but with one of double or triple the period.

In Figure 6, we show two periodic solutions corresponding to a forcing term $\lambda \sin \mu t$ with $\mu = 2.6$. The solution is similar to that found at $\mu = 1.3$. The figure looks a little different because the last sixteen periods of the forcing term are shown so the x-scale is half the one shown in Figure 4.

All the remarks of the previous section apply to solutions that are subharmonic responses.

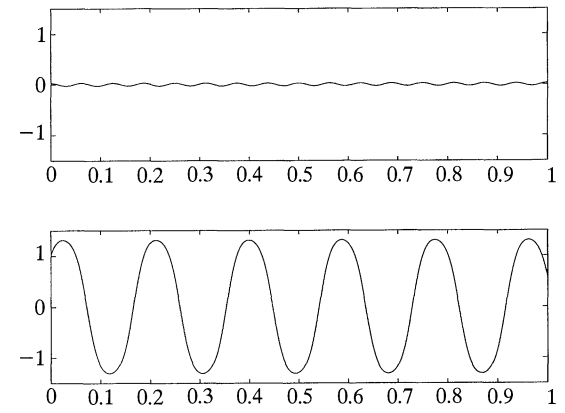


Figure 6. A subharmonic response to periodic forcing. Here, with $\mu = 2.6$, the large and small initial conditions give rise to either a small linear response of the same period or to one of twice the period similar to that seen around $\mu = 1.3$.

The most intriguing values of μ to investigate may be when it is near 4. This is because the *vertical* motions that preceded the torsional ones had a frequency (about 40 per minute) that corresponded to approximately this value, strongly suggesting the existence of some forcing term of this frequency.

Sure enough, when the nonlinear system is forced at this frequency, one obtains a periodic response to a large initial condition and a small forcing term of approximately the right frequency and magnitude. This is shown in Figure 7, where a forcing term $\lambda \sin 4t$ is taken, and combined with either a large or small initial condition. Varying the initial conditions can result in either small linear responses of the right period (with small initial conditions) or large nonlinear responses similar to those described earlier, with period three times that of the forcing term.

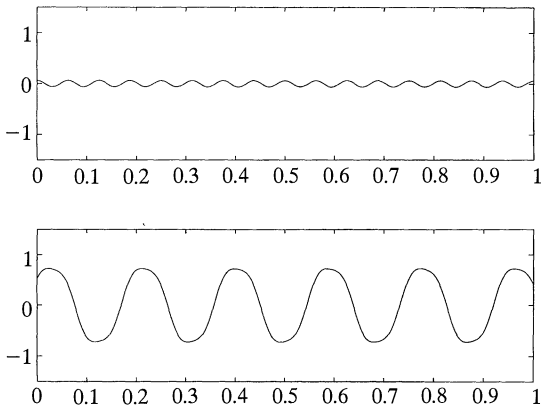


Figure 7. These torsional oscillations corresponds to large and small initial conditions if $\mu = 4$, corresponding to the frequency of the earlier vertical oscillations in the bridge, before the onset of torsional motion.

3.3 Some transient results. Now we illustrate how dramatically different the transient behaviour is for the trigonometrically correct and the linearized system. We solve the initial value problem for a variety of different initial conditions and compare the two systems. Throughout this subsection, where a forcing term has been used, it is of the form $\lambda \sin(1.2t)$, although similar results would occur around $\mu = 1.3$ or 1.4.

First, the good news.

If we solve the initial value problem with *no forcing* but with large initial conditions, the results over the first hundred periods for the two problems are shown in Figure 8. There is little real difference, as the damping takes over and both systems settle back to equilibrium.

Now, do the same experiment starting with initial conditions at equilibrium, but with a small forcing term $\lambda = 0.05$. Again, the two systems give close results. This is shown in Figure 9.

So far, so good! When subjected separately either to small periodic forcing or to a large transient displacement from equilibrium, the system responds with much the same behaviour in the linear and nonlinear models.

Now to the results that we have already discussed. If we combine the previous two effects, the principle of superposition predicts that the linear system will die down as before. Figure 10 shows this, as well as the huge difference caused by doing the correct trigonometry. The trigonometrically correct model continues

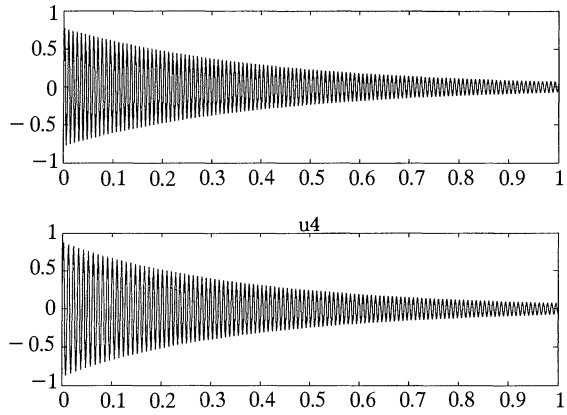


Figure 8. The result of a large push from equilibrium, in both the linear and correct models, with no forcing term.

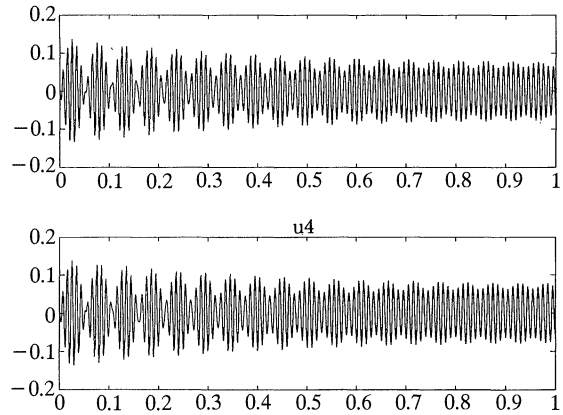


Figure 9. The result of a small forcing term starting at equilibrium. Both systems start to die down immediately.

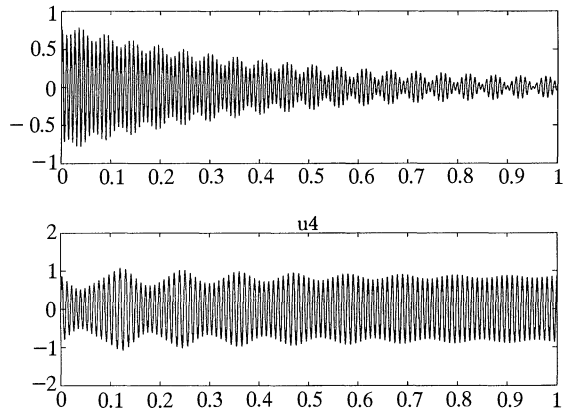


Figure 10. The transient behaviour resulting from combining the two influences shown separately in Figures 8 and 9. Predictably, the linear model starts to die down but the nonlinear one goes into large oscillation.

indefinitely in a large torsional oscillation, eventually settling to the large periodic motion introduced in Figure 2.

Finally, we show how sensitive the correct system is to the amplitude of the forcing term. Figure 11 shows the effect of changing the amplitude of the forcing term from $\lambda = 0.05$ (top), which results in the large amplitude motion, to $\lambda = 0.04$ (bottom), when the motion begins immediately to die down.

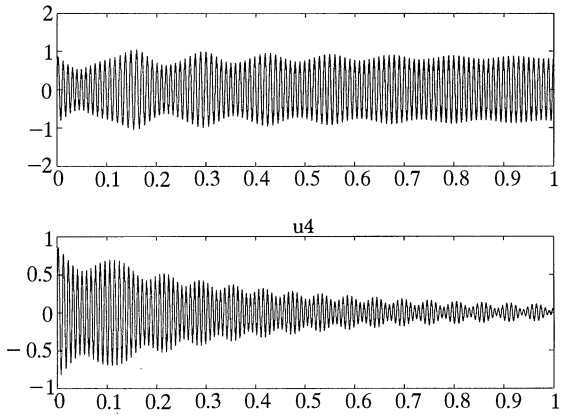


Figure 11. A slight change in the amplitude of the forcing term can make a huge difference in the nonlinear model. The effects of a large push with $\lambda = 0.04$ and $\lambda = 0.05$ are compared. The slight increase is enough to change the eventual behaviour from small near-equilibrium motion to large and destructive motion.

3.4 A quick summary. Performing the unnatural act of linearizing the trigonometric oscillator of equation (10) has the effect of removing a large class of large amplitude behaviours.

As shown in Figures 8 and 9, the linear model is good in the presence of a large transient push and no forcing, or if one starts off at equilibrium with small forcing. However, when these two effects are combined, it completely fails to predict the eventual long-term behaviour of the real system (10).

In the trigonometrically correct model, over a wide range of frequency, several different amplitude oscillations can exist for the same periodic forcing term. Which one the system ends up on can depend on simple transient events, such as a single large push. Large amplitude solutions can result as a subharmonic response to a higher frequency forcing term and are of roughly the right magnitude and frequency when compared to the oscillation of the Tacoma Narrows bridge in its final torsional oscillations.

We emphasize that there is nothing high-tech or controversial about these results.

It is well known that a periodically forced pendulum equation such as (5) exhibits multiple solution and chaotic behaviour [**10**]. Perhaps the only surprise is that when one makes reasonable guesses for the constants based on the available literature, the resulting periodic motions are so good a match to the historical data.

One simply recognizes that trigonometry has been artificially removed from the problem, restores it, and uses a good Runge-Kutta initial value solver to find the long-time behaviour of the system.

There remain two other things to understand: first, what was the origin of the original vertical oscillations of up to 5 ft. in amplitude and with a frequency of about 40 per minute? If we assume these oscillations were approximately sinusoidal, accelerations must have been approaching that of gravity.

Second, one can ask about the origin of the sudden transition from vertical to torsional motion. On this question, we may need to be more speculative or controversial, as we shall see in the next section. However, as regards the large amplitude torsional periodic solution, we believe the explanation of this section is satisfactory.

4. WHAT IF THE SPRINGS LOSE TENSION BRIEFLY? THE TRANSITION TO TORSIONAL MOTION. So far, we have focussed on the least controversial results: namely, if one does not artificially remove the trigonometry from the torsional oscillator, one gets a realistic explanation for the oscillations seen in the Tacoma Narrows bridge before its collapse. We have largely avoided the question of what happens when the two equations (3) and (4) are coupled due to periodic slackening of the cables.

If this does happen, the structure of the periodic solutions becomes considerably more complex. A start to studying this system and hints of the complexities that arise can be found in [**12**].

There is some controversy about this question. I and co-workers claim that with accelerations reaching that of gravity, and with magnitudes of up to 28 ft. every four seconds, there must have been some slackening of the cables. Some engineers insist, to the verge of apoplexy, that this cannot happen [**3**]. A third group take a middle path, claiming that the cables can slacken only in a brief and transitory fashion [**3**].

In this final section, we follow the middle path and ask what happens to the initial value problem solution of the full nonlinear system, if the cables *briefly* lose tension due to a vertical motion.

Here, we come to a new and unexpected conclusion: *a purely vertical nontorsional motion in which the cables lose tension can become disastrously unstable even in the presence of tiny torsional forces, setting up a rapid transition to large amplitude torsional motion.* This might be the explanation for the sudden transition observed at Tacoma Narrows.

We have no good mathematical explanation for this phenomenon. However, we can demonstrate it in a sequence of mathematical experiments.

We start with the initial value problem (3) and (4), with a fairly large initial impulse in the vertical direction and no torsional oscillation. Figure 12 shows the effect of a large push in the vertical direction, $y(0) = 26$ with *tiny* torsional forcing ($\lambda = 0.002$). What you get is exactly what you expect: the cables loosen briefly for about two periods (falling below $y = 0$), but the torsional motions are confined to about $\pm$ 0.003, essentially undetectable.

Now repeat the same experiment, changing only the magnitude of the initial push in the vertical direction, to $y(0) = 31$. The results are shown in Figure 13. The vertical y-motion (top) is behaving much as before, but the character of the torsional motion has totally changed. The additional vertical push has resulted in a completely different torsional oscillation. There is initially almost none, and then, virtually instantaneously, it changes to large amplitude torsional oscillations of magnitude approaching 1 radian. Since there was essentially no torsional forcing term, damping eventually takes over and the motion settles down to the near

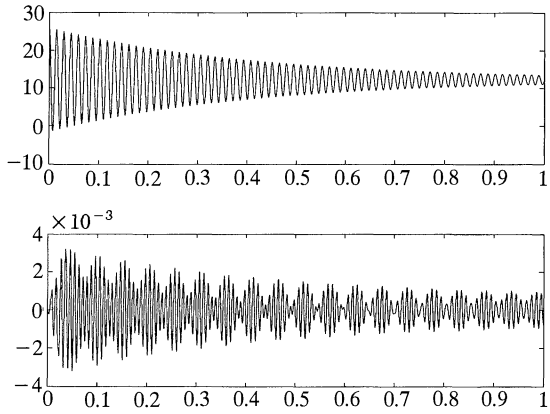


Figure 12. Vertical (top) and torsional (bottom) responses to a large push in the vertical direction. As intuition would suggest, the vertical oscillation is large but settles down quickly. Torsional oscillation with only a tiny forcing term is essentially invisible.

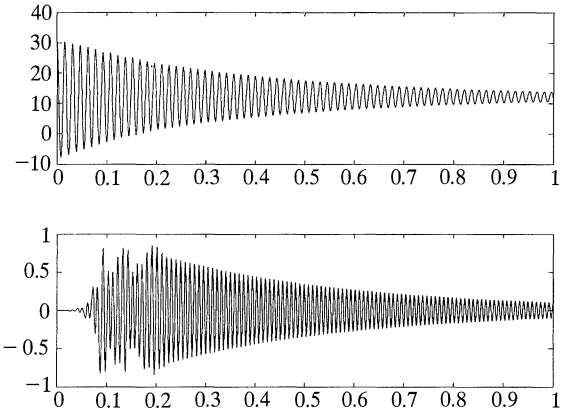


Figure 13. A somewhat larger initial push in the vertical direction, but with everything else the same as in Figure 12, gives rise to a sudden transition to large amplitude torsional oscillation.

equilibrium behaviour. Of course, if there had been no torsional forcing, as opposed to tiny, there would have been no torsional result.

Suppose we have no torsional forcing term. Then if $\theta(0) = 0$, the result is pure vertical motion. Now, repeat the same experiment with $\theta(0) = 0.001$. The result is shown in Figure 14: nothing torsional for a while, followed by a sudden jump to violent torsional (up to 50 deg), followed by eventual decay to equilibrium as the damping takes over.

We conjecture that this might have been the situation at the instant of changeover at Tacoma Narrows from vertical to torsional oscillation, which appeared to take place almost instantaneously. There was already a large amount of vertical kinetic energy built up and a small torsional forcing term could have been enough to change the motion.

The reader probably can guess where we are heading. Suppose we have a small amount of periodic torsional forcing, enough to sustain a ± 3 deg motion near

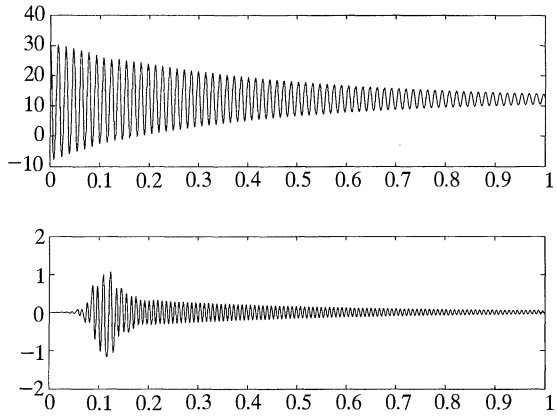


Figure 14. No torsional forcing term, but a tiny displacement in the initial value of $\theta(0)$ is also enough to induce violent torsional oscillation, which, in the absence of periodic forcing, eventually dies down due to damping.

equilibrium. We have already seen how a single large *torsional* push could induce a large-amplitude periodic torsional motion. Now we consider the influence of a large *vertical* push combined with the small amount of periodic torsional forcing. The result is shown is Figure 15.

Here we have taken μ as 1.3, in the middle of the range where we expect multiple solutions. We have taken $\lambda = 0.02$, about enough to induce linear oscillation of about ± 2 deg. We have given a sufficiently large push in the vertical direction to cause the springs to lose tension for the first five cycles. The result is a couple of cycles of small torsion, followed by an almost instantaneous transition to huge torsional motion, which remains permanent. It eventually settles down to the permanent periodic torsional motion shown in Figure 4.

We believe that we have discovered a convincing explanation for the mystery of the sudden transition to torsional motion. *A large vertical motion had built up, there*

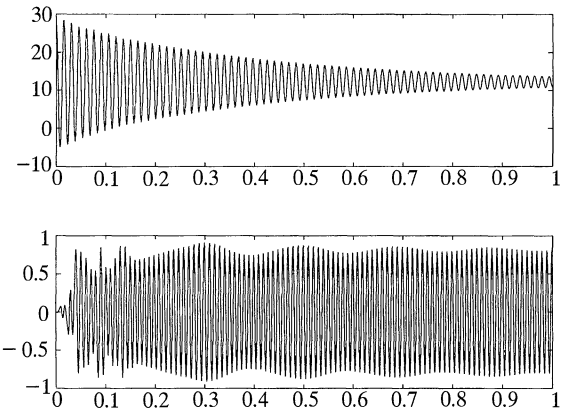


Figure 15. What really happened to induce torsional oscillation at Tacoma Narrows? A small but periodic torsional force combines with a large vertical transient push to produce a rapid transition to large torsional oscillation as the vertical oscillation is damped away.

was a small push in the torsional direction to break symmetry, the instability occurred, and small aerodynamic torsional periodic forces were sufficient to maintain the large periodic torsional motions, as shown in Figure 15.

5. SOME CONCLUDING COMMENTS. No mathematical model is ever perfect. Turing said it best: "This model will be a simplification and an idealization, and consequently a falsification. It is to be hoped that the features retained for discussion are those of the greatest importance in the present state of knowledge" [**32**]. Let us review some of the short-comings of our paper.

Following the engineering literature, we have treated the cable-suspension structure as a torsional oscillator supported by springs that remain linear until they reach the unloaded state. We doubt that a bridge oscillating up and down by about 10 meters every 4 seconds obeys Hooke's law.

Our model slightly understates the period of the large amplitude torsional oscillations. This is probably due to the fact that there is additional resistance to torsion from the road-bed, adding an extra spring constant to equation (10).

Our model says very little about the vertical oscillation that preceded the torsional oscillation. It may explain why the vertical motion was so rapidly converted to torsional motion, but has little to say about why this original motion started and continued. Nor does it say much about the rather mysterious complex torsional motion that actually occurred, namely one that alternated between one-noded and no-noded oscillations.

However, it still gives a remarkably "low-tech" explanation of two of the phenomena, the large amplitude torsional motions and the transitional motions.

There remains a great deal to do on the coupled system. With a reliable Runge-Kutta solver and unlimited computer time (and some patience), the reader can discover new phenomena in the solution set of this system. Certainly, this type of experimentation makes for interesting class projects in the undergraduate environment.

A mathematical explanation for the apparently unstable nature of the large amplitude vertical oscillation in which cables lose tension would be desirable. Although large amplitude vertical oscillations have been investigated and their one-dimensional stability proved [**13**], we have no proof of their instability in the torsional direction. An intuitive argument might be advanced that if the rod is ever in the situation where one spring is under tension and the other is not, this introduces a new large torsional force.

Finally, we are not sure of the consequences of this work for modern suspension bridges in earthquake conditions. Part of the dilemma, as one leading bridge engineer has lamented, is that there is a "lack of open discussion" on these problems [**6**].

It is not clear whether, in their calculations about earthquake responses, engineers have taken into account the potentially catastrophic consequences of a brief loosening of the cables and the ensuing large amplitude torsional oscillations that can result. To judge by the literature, they are still making the small oscillation linearization, which can be so misleading once away from equilibrium [**1**]. Since hundreds of millions of dollars are being spent in an effort to strengthen suspension bridges in California in preparation for large earthquakes, this question may not be entirely academic.

Finally, it is worth remarking that our results illustrate how the availability of inexpensive computation is changing the entire culture of mathematics. Now, new

and interesting results can be discovered in the undergraduate classroom. We can now investigate numerically simple systems such as (4) and (5), and we uncover beautiful new properties that we could not have suspected previously. We may be witnessing the dawn of a new golden age of discovery in nonlinear oscillations.

REFERENCES

1. A. M. Abdel-Ghaffar, Suspension bridge vibration: Continuum formulation, *Jour. Eng. Mechanics, A.S.C.E.* **108** (1982) 1215–1232.
2. O. H. Amann, T. von Kármán, and G. B. Woodruff, *The Failure of the Tacoma Narrows Bridge*, Federal Works Agency, 1941.
3. D. Berreby, The great bridge controversy, *Discover* **13** (1992) 26–33.
4. K. Y. Billah, and R. H. Scanlan, Resonance, Tacoma Narrows bridge failure, and undergraduate physics textbooks, *Amer. J. Physics* **59** (1991) 118–124.
5. F. Bleich, C. B. McCullough, R. Rosecrans, and G. S. Vincent, *The Mathematical Theory of Suspension Bridges*. U.S. Dept. of Commerce, Bureau of Public Roads, 1950.
6. A. Castellani, Safety Margins of suspension bridges under seismic conditions, *ASCE J. Structural Engineering* **113** (1987) 1600–1616.
7. A. Castellani, and P. A. Felotti, A note on lateral vibration of suspension bridges, *ASCE J. Structural Engineering* **112** (1986) 2169–2173.
8. Y. Chen, and P. J. McKenna, Traveling waves in a nonlinearly suspended beam: theoretical results and numerical observations, *J. Differential Equations*, **136** (1997) 325–355.
9. Y. S. Choi, K. C. Jen, and P. J. McKenna, The structures of the solution set for periodic oscillations in a suspension bridge model, *IMA J. Appl. Math.* **47** (1991) 283–306.
10. P. Blanchard, R. L. Devaney, and G. Hall, *Differential Equations*, PWS Publishing, Boston, 1996.
11. S. H. Doole, and S. J. Hogan, A piecewise linear suspension bridge model: Nonlinear dynamics and orbit continuation, *Dynamics Stability Systems* **11** (1996) 19–47.
12. S. H. Doole, and S. J. Hogan, *Torsional dynamics in a simple suspension bridge model*, Applied Nonlinear Mathematics Research Report Number 9.96, University of Bristol, Bristol, 1996.
13. J. Glover, A. C. Lazer, and P. J. McKenna, Existence and stability of large-scale nonlinear oscillations in suspension bridges, *Z. Angew. Math. Phys.* **40** (1989) 171–200.
14. D. Jacover and P. J. McKenna, Nonlinear torsional flexings in a periodically forced suspended beam, *J. Comput. Appl. Math.* 52 (1994) 241–265.
15. Theodore von Kármán, *The Wind and Beyond, Theodore von Kármán, Pioneer in Aviation and Pathfinder in Space*, Little Brown and Co. Boston, 1967.
16. A. C. Lazer, and P. J. McKenna, Large amplitude periodic oscillations in suspension bridges: some new connections with nonlinear analysis, *SIAM Review* **32** (1990) 537–578.
17. A. C. Lazer, and P. J. McKenna, Large scale oscillatory behaviour in loaded asymmetric systems, *Analyse Nonlinéaire Annales de L'Institut Henri Poincare* **4** (1987) 243–274.
18. P. J. McKenna, and W. Walter, Nonlinear oscillations in a suspension bridge, *Arch. Rat. Mech. Anal.* **98** (1987) 167–177.
19. P. J. McKenna, and W. Walter, Travelling waves in a suspension bridge, *SIAM J. Appl. Math.* 50 (1990) 703–15.
20. I. Peterson, Rock and Roll Bridge, *Science News* **137** (1990) 344–346.
21. H. Petroski, Still twisting, *American Scientist*, September–October, 1991.
22. Mario Salvadori, personal communication.
23. R. H. Scanlan, Developments in low-speed aeroelasticity in the civil engineering field, *AIAA Journal* **20** (1982) 839–844.
24. R. H. Scanlan, Airfoil and bridge deck flutter derivatives, *Proc. Amer. Soc. Civ. Eng. Eng. Mech. Division* EM6 (1971) 1717–1737.
25. R. H. Scanlan, The action of flexible bridges under wind II: buffeting theory., *J. Sound and Vibrations* **60** (1978) 201–211.
26. R. H. Scanlan, and J. J. Tomko, Airfoil and bridge deck flutter derivatives, *Proc. Amer. Soc. Civ. Eng. Eng. Mech. Division* EM6 (1971) 1717–1737.

27. R. H. Scanlan, and J. W. Vellozi, Catastrophic and annoying responses of long-span bridges to wind action, *Annals New York Acad. Sciences* **352** (1980) 247–263.
28. F. D. Schwarz, Why theories fall down, *American Heritage of Inventions and Technology* **8** (1993) 6–7.
29. F. D. Schwarz, Still Falling, *American Heritage of Inventions and Technology* **9** (1993) 7.
30. Seattle Times/Seattle Post-Intelligencer, November 5, 1990 p. 12.
31. J. L. Synge, and B. A. Griffith, *Principles of Mechanics*, McGraw-Hill, New York, 1959.
32. A. M. Turing, The chemical basis of morphogenesis, *Philos. Trans. Royal Soc.* Ser. B 237 (1952) 37–72.
33. P. Vielsack, and H. Wei, Sensivity of the harmonic oscillation of a suspension bridge model with asymmetric dissipation, *Arch. Appl. Mech.–Ingenieur Archiv.* **64** (1994) 408–416.

P. J. MCKENNA did his undergraduate work in Dublin (at U.C.D.) and his graduate work in Ann Arbor. The central theme of his research is nonlinear analysis, in particular, the existence, multiplicity, and numerical approximation of solutions of nonlinear boundary value problems. The work arises naturally from his research in multiple periodic solutions of Hamiltonian systems. Kristen Moore, his doctoral student, is extending this analysis to the partial differential equations that describe the spatial behaviour along the length of the bridge. Movies showing computed solutions can be seen at `http:\\www.math.uconn.edu\~kmoore\`
University of Connecticut U-9, Storrs, CT 06269
mckenna@math.uconn.edu

Cicero on mathematics ....

For indeed you cannot fail to remember that the most learned men hold what the Greeks call 'philosophy' to be the creator and mother, as it were, of all the reputable arts, and yet in this field of philosophy it is difficult to count how many men there have been, eminent for their learning and for the variety and extent of their studies, men whose efforts were devoted, not to one separate branch of study, but who have mastered everything they could whether by scientific investigation or by methods of dialectic. Who does not know, as regards the so-called mathematicians, what very obscure subjects, and how abstruse, manifold, and exact an art they are engaged in? Yet in this pursuit so many men have displayed outstanding excellence, that hardly one seems to have worked in real earnest at this branch of knowledge without attaining the object of his desire. Who has devoted himself wholly to the cult of the Muses, or to this study of literature, which is professed by those who are known as men of letters, without bringing within the compass of his knowledge and observation the almost boundless range and subject-matter of those arts?

De Oratore, I. iii. 9–10
Contributed by Adi Ben-Israel, Rutgers University

Inverse Conjugacies and Reversing Symmetry Groups

Geoffrey R. Goodson

INTRODUCTION. We present some elementary group theory that arises in the theory of time-reversing dynamical systems. Let G be a (usually non-abelian) group and let $a \in G$ be a fixed element. The set

$$C(a) = \{x \in G : xa = ax\},$$

the *centralizer* of a, is a subgroup of G that contains $\langle a \rangle$, the cyclic subgroup generated by a. Our aim is to study the *skew centralizer*

$$B(a) = \{x \in G : xa = a^{-1}x\}.$$

This paper arose from a course I gave on algebraic structures, where some of the results of Sections 1 and 2 and some examples from Section 4 were presented as exercises and then discussed in the classroom. In addition, the students were asked to calculate $B(a)$ and $C(a)$ for certain specific examples, sometimes with the aid of a software package.

Generally $B(a)$ is not a subgroup of G, and it may be empty. However, $E(a) = B(a) \cup C(a)$ is a group, which is called the *reversing symmetry group* of a. In dynamical systems theory, the group element a represents the time evolution operator of a dynamical system. We present some results familiar to people working in time reversing dynamical systems, but our presentation is given in an abstract setting, entirely from an elementary group theoretic point of view, in the hope that it will be of interest to teachers of a first course in group theory.

Section 1 gives some of the elementary properties of $B(a)$. We see that $B(a)$ is a group if and only if a is an involution, i.e., $a^2 = e$, the identity of G. In Section 2, we prove (following Lamb [**8**]) that $E(a)$ is a group having $C(a)$ as a normal subgroup. In dynamical systems, the case where a has infinite order is of most interest, but we show that there are interesting finite groups, such as the dihedral and dicyclic groups, that arise in a natural way from the study of $E(a)$. In Section 3 we study the inner automorphisms of $E(a)$ and apply them when G is a topological group. We give particular emphasis to the situation when $\{s^2 : s \in B(a)\}$ is a singleton set. Inverse conjugacies involving the permutation groups and some infinite groups originating in dynamical systems theory are our focus in Section 4.

In Section 5 we mention briefly the dynamical origins of the ideas discussed here, restricting our attention to the ergodic theory of measure-preserving transformations.

1. ELEMENTARY PROPERTIES OF $B(a)$. Elements $a, b \in G$ are said to lie in the same *conjugacy class* (that is, a and b are *conjugate*) if there exists some $x \in G$ satisfying $a = x^{-1}bx$. Thus $a \in G$ is conjugate to its inverse if $B(a) \neq \emptyset$.

The set $B(a)$ is a group only in special circumstances. If it is a group, then $B(a)$ contains the identity of G, which we denote by e, so $ea = a^{-1}e$, $a = a^{-1}$ or $a^2 = e$.

This in turn implies that $B(a) = \{x \in G : xa = a^{-1}x\} = C(a)$. On the other hand, if $B(a) = C(a)$, then $B(a)$ is a group. Let us also mention that if $B(a) \cap C(a)$ is non-empty, then we must have $a^2 = e$. We have proved:

Proposition 1. *For a group G and a given $a \in G$ with $B(a) \neq \emptyset$, the following are equivalent*:

(*i*) *$B(a)$ is a subgroup of G.*
(*ii*) *a is an involution, i.e., $a^2 = e$.*
(*iii*) *$B(a) = C(a)$.*
(*iv*) *$B(a) \cap C(a) \neq \emptyset$.*

If G is an abelian group, then $C(a) = G$, so from Proposition 1, either $B(a) = C(a)$ (when $a^2 = e$), or $B(a) = \emptyset$. If any of the conditions of Proposition 1 hold, we say that we have the *trivial case*.

In general, $B(a)$ may have elements of all even orders, and also of infinite order. However, if there exists an $x \in B(a)$ with $x^n = e$ for some odd $n \in \mathbb{Z}$, then

$$xa = a^{-1}x \Rightarrow x^n a = a^{-1}x^n \Rightarrow a = a^{-1} \Rightarrow a^2 = e,$$

so we again have the trivial case.

On the other hand, suppose $x \in B(a)$ is of order $2n$ and 2^m is the highest power of 2 dividing $2n$, i.e., $2n = 2^m k$ for some odd $k \in \mathbb{Z}$. Then $y = x^k \in B(a)$ and $y^{2^m} = x^{k \cdot 2^m} = x^{2n} = e$, so $B(a)$ also contains elements of order 2^m (see [**9**, p. 14]).

Notice that $a \in G$ is conjugate to a^{-1} if and only if there are $u, v \in G$ such that

$$a = vu^{-1} \quad \text{and} \quad u^2 = v^2.$$

To prove this, take any w such that $aw = wa^{-1}$. Then $(aw)^2 = awaw = wa^{-1}aw = w^2$. Set $v = aw \in B(a)$ and $u = w$. Then $a = vu^{-1}$ and $v^2 = (aw)^2 = w^2 = u^2$. Conversely, suppose $a = vu^{-1}$ and $u^2 = v^2$. Then $au = vu^{-1}u = v$ and $ua^{-1} = u(vu^{-1})^{-1} = u^2v^{-1} = v^2v^{-1} = v$, i.e., $au = ua^{-1}$.

The case where there exists an involution in $B(a)$ is important in the dynamical systems literature. We now see that a is conjugate to a^{-1} via an involution if and only if there are $u, v \in G$ such that $a = uv^{-1}$ and $u^2 = v^2 = e$ (see [**9**, p. 13] and [**4**]).

2. THE REVERSING SYMMETRY GROUP $E(a) = B(a) \cup C(a)$. We claim that the set $E(a) = B(a) \cup C(a)$, is a subgroup of G. This is clear if $B(a)$ is empty, so assume it is non-empty. Taking inverses of both sides of $xa = a^{-1}x$ gives $a^{-1}x^{-1} = x^{-1}a$, so

$$x \in B(a) \Leftrightarrow x^{-1} \in B(a) \quad \text{and} \quad xa = a^{-1}x \Leftrightarrow ax = xa^{-1}.$$

Since $B(a)$ and $C(a)$ are closed under the taking of inverses, so is $E(a)$.

Let $x, y \in E(a)$. If $x, y \in C(a)$, then $axy = xay = xya$. Therefore $xy \in C(a)$ and so $xy \in E(a)$.

If $x, y \in B(a)$, then $axy = xa^{-1}y = xya$. Therefore $xy \in C(a)$ and again $xy \in E(a)$.

The third possibility is $x \in B(a)$, $y \in C(a)$. In this case $axy = xa^{-1}y = xya^{-1}$, so $xy \in B(a)$, and $xy \in E(a)$.

The proof that $E(a)$ is a group is completed on noting that $e \in C(a) \subseteq E(a)$.

The preceding argument shows that $C(a)$ is a subgroup of $E(a)$. Suppose $a^2 \neq e$ and $B(a) \neq \emptyset$, and let $x, y \in B(a)$. Then $x^{-1}y \in C(a)$ and $y \in x \cdot C(a)$. It follows that $B(a) \subseteq x \cdot C(a)$ and in a similar way that $x \cdot C(a) \subseteq B(a)$. In particular, the cosets of $C(a)$ in $E(a)$ are $C(a)$ and $x \cdot C(a) = B(a)$.

Clearly $x \cdot C(a) \cdot x^{-1} = C(a)$ for all $x \in E(a)$, so $C(a)$ is a normal subgroup of $E(a)$, and we see that $E(a)/C(a) \cong \mathbb{Z}_2$. If $\langle a \rangle = \{a^n : n \in \mathbb{Z}\}$ is the cyclic subgroup generated by a, then $\langle a \rangle$ is a normal subgroup of both $C(a)$ and $E(a)$.

If G is a finite group having even order, then it is an easy exercise to show that there exists some $a \in G$, $a \neq e$, with $a^2 = e$; the conditions of Proposition 1 are satisfied for a, so $B(a) = C(a) \neq \emptyset$. Consequently, finite groups of even order always contain nontrivial elements that are conjugate to their inverse.

On the other hand, finite groups of *odd* order never contain a nontrivial element that is conjugate to its inverse. Suppose the order of G is odd and there is an $a \in G$, $a \neq e$, that is conjugate to its inverse. Lagrange's theorem implies that a cannot be of even order, so $a^2 \neq e$. By assumption, $B(a) \neq \emptyset$, so Proposition 1 implies that $E(a) = B(a) \cup C(a)$ is a disjoint union. Since the cosets of $C(a)$ in $E(a)$ are $C(a)$ and $B(a)$, $E(a)$ is a subgroup of G of even order. Lagrange's Theorem again tells us that this is impossible.

The fact that $E(a)$ is a group with $E(a)/C(a) \cong \mathbb{Z}_2$ appears in [**8**] and [**9**, p. 9–12]. However, the following may not be well known:

Proposition 2. *Let $a \in G$ with $B(a) \neq \emptyset$.*

(*i*) *If $x \in B(a)$ and $x^2 \in \langle a \rangle$, then $x^4 = e$. In particular, if the order of $\langle a \rangle$ is infinite or odd, then $x^2 = e$.*

(*ii*) *If $C(a) = \langle a \rangle$, then $\{x^2 : x \in B(a)\}$ is a singleton subset of $C(a)$.*

(*iii*) *If $\{x^2 : x \in B(a)\}$ is a singleton set, then the order of x divides 4 for all $x \in B(a)$.*

(*iv*) *The center of $E(a)$ is a subgroup of $C(a)$.*

Proof: (i) If $x \in B(a)$ and $x^2 \in \langle a \rangle$, then $x^2 = a^n$ for some $n \in \mathbb{Z}$. Since

$$xa = a^{-1}x \Rightarrow xa^n = a^{-n}x \Rightarrow x \cdot x^2 = x^{-2} \cdot x \Rightarrow x^4 = e,$$

we see that the order of x divides 4 and hence the order of x^2 divides 2. If $x^2 \in \langle a \rangle$, a cyclic group having infinite (or odd) order, then $\langle a \rangle$ cannot have any elements of order 2. Therefore $x^2 = e$.

(ii) In a similar way, if $x_1, x_2 \in B(a)$, then $x_1x_2 \in C(a) = \langle a \rangle$, so $x_1x_2 = a^n$ for some $n \in \mathbb{Z}$. Also $x_1a^n = a^{-n}x_1$, so $x_1x_1x_2 = (x_1x_2)^{-1}x_1$, or $x_1^2 = x_2^{-2} = x_2^2$, so $\{x^2 : x \in B(a)\}$ is a singleton set.

(iii) If $x \in B(a)$, then $x^3 \in B(a)$ so $x^2 = (x^3)^2 = x^6$, and this implies $x^4 = e$.

(iv) If x is in the center of $E(a)$, then $xg = gx$ for all $g \in E(a)$. In particular, $xa = ax$, so $x \in C(a)$. ∎

Since $\langle a \rangle$ is a normal subgroup of $C(a)$, the quotient group $C(a)/\langle a \rangle$ is well defined. This group is sometimes called the *essential centralizer* of a. Proposition 2(ii) can be generalized to:

If $a \in G$ has infinite order and $C(a)/\langle a \rangle$ has order m for some $m \in \mathbb{Z}^+$, then every $x \in B(a)$ has order $2k$ for some k that divides m.

3. THE INNER AUTOMORPHISMS OF $E(a)$. The inner automorphisms of $E(a)$ have the form $\phi_s(x) = sxs^{-1}$ for some $s \in E(a)$. Since $C(a)$ and $\langle a \rangle$ are normal subgroups of $E(a)$, they are preserved by ϕ_s for all $s \in E(a)$. Furthermore, if $s \in B(a)$ and $x \in \langle a \rangle$, then $\phi_s(x) = x^{-1}$. This is because $x = a^n$ for some $n \in \mathbb{Z}$, and so $\phi_s(x) = sa^n s^{-1} = a^{-n}ss^{-1} = a^{-n} = x^{-1}$.

When is it true that $\phi_s(x) = x^{-1}$ for all $x \in C(a)$? Let us say that $s \in B(a)$ *conjugates* $C(a)$ *to* $C(a)^{-1}$ if $sx = x^{-1}s$ for all $x \in C(a)$, or, equivalently, if $\phi_s(x) = x^{-1}$ for all $x \in C(a)$.

We can now show that every $s \in B(a)$ conjugates $C(a)$ to $C(a)^{-1}$ if and only if $\{s^2 : s \in B(a)\}$ is a singleton set.

To see this we use the fact that $B(a) = sC(a)$ for each $s \in B(a)$. If $s \in B(a)$ conjugates $C(a)$ to $C(a)^{-1}$ then $sx = x^{-1}s$ for all $x \in C(a)$. This implies that $(sx)^2 = s^2$, and the result follows.

Conversely, suppose that $\{s^2 : s \in B(a)\}$ is a singleton set. Then $(sx)^2 = s^2$ for any $s \in B(a)$ and $x \in C(a)$. This immediately gives $sx = x^{-1}s$, so s conjugates $C(a)$ to $C(a)^{-1}$.

We leave it to the reader to show that if $\{s^2 : s \in B(a)\}$ is a singleton set, then $C(a)$ is abelian.

Our aim now is to apply the preceding results to the case where G is a metrizable topological group. Note that if $\langle a \rangle$ is dense in $C(a)$, then $C(a)$ must be abelian, and either $C(a) = \langle a \rangle$ or $C(a)$ is uncountable. An immediate consequence of the next theorem is that if $\langle a \rangle$ is dense in $C(a)$ and $B(a) \neq \varnothing$, then every $s \in B(a)$ conjugates $C(a)$ to $C(a)^{-1}$.

Theorem 1. *If $\langle a \rangle$ is dense in $C(a)$ and $B(a) \neq \varnothing$, then $\{s^2 : s \in B(a)\}$ is a singleton set.*

Proof: We are given that $\overline{\{a^n : n \in \mathbb{Z}\}} = C(a)$. Let $x_1, x_2 \in B(a)$, so $x_1x_2 \in C(a)$. We use an argument similar to that in Proposition 2(ii).

There is a subsequence $\{n_i\}$ of integers for which $x_1x_2 = \lim_{i\to\infty} a^{n_i}$. Furthermore, $ax_1 = x_1a^{-1}$ implies that $a^{n_i}x_1 = x_1a^{-n_i}$ for all i.

Letting $i \to \infty$ and using the continuity of multiplication and inversion in G, we obtain

$$x_1x_2x_1 = x_1(x_1x_2)^{-1} = x_1x_2^{-1}x_1^{-1} \Rightarrow x_2x_1 = x_2^{-1}x_1^{-1} \Rightarrow x_2^2 = x_1^{-2}.$$

The fact that $B(a)$ is closed under the taking of inverses now gives $x_1^2 = x_2^2$, so $\{x^2 : x \in B(a)\}$ is a singleton set. ∎

If $x \in B(a)$, then we cannot have $a \in \{x^{2n} : n \in \mathbb{Z}\}$ as this would imply $a^2 = e$, the trivial case. Whenever G is a topological group, the set $\{x^{2n} : n \in \mathbb{Z}\} \subseteq C(a)$ is never dense in $C(a)$ for any $x \in B(a)$.

This section is based on [**3**] and [**4**]; see also [**10**], where inner automorphisms on $E'(a)$, the *reversing k–symmetry group* are discussed.

4. EXAMPLES.

(i) Dihedral and Dicyclic Groups. The *dihedral group* D_n of order $2n$ arising from the symmetries of a regular plane n-gon is

$$\langle a, x : a^n = e,\ x^2 = e,\ x^{-1}ax = a^{-1} \rangle.$$

The *dicyclic groups* of order $4m$ are

$$\langle a, x : a^{2m} = e,\ x^2 = a^m,\ x^{-1}ax = a^{-1} \rangle.$$

See [**1**, p. 6–8] or [**12**, p. 65–66] for a discussion of these groups.

If a has finite order n and $C(a) = \langle a \rangle$, Proposition 2(ii) says that $x^4 = e$ if $x \in B(a)$. Essentially two different cases arise:

(a) If $x^2 = e$, then $E(a) \cong D_n$.
(b) If $x^2 \neq e$, then $x^2 = a^{n/2}$. In particular, n is even and $E(a)$ is a dicyclic group of order $2n$.

As a concrete example, consider the smallest of the dicyclic groups, the quarternion group $Q_8 = \{\pm 1, \pm i, \pm j, \pm k\}$ with multiplication given by $i^2 = j^2 = k^2 = -1$, $ij = k$, $jk = i$, $ki = j$, and $ji = -k$, $kj = -i$, $ik = -j$ and the usual rules for multiplying by ± 1. Then $C(i) = \{\pm 1, \pm i\}$ and $B(i) = \{\pm j, \pm k\}$. Since $C(i) = \langle i \rangle$ is a cyclic group, $\{x^2 : x \in B(i)\} = \{-1\}$ is a singleton set.

(ii) Permutation Groups. Every element of the permutation group S_n on n symbols is conjugate to its inverse. This is because any permutation in S_n can be resolved into a product of disjoint cycles in a unique manner except for the order in which the cycles appear. Two permutations are in the same conjugacy class if they have the same cycle pattern. For example, in S_8 the permutations

$$x = (3,5,6)(2,4)(7,1) \quad \text{and} \quad y = (2,4,1)(3,5)(6,7)$$

have the same cycle pattern, so they lie in the same conjugacy class. It is clear that every permutation and its inverse have the same cycle pattern.

Not every element of the alternating group A_n need be conjugate to its inverse. For example in A_4, the cycles $(1,2,3)$ and $(1,2,3)^{-1}$ are not conjugate. This is because $B((1,2,3))$ (in S_4) consists solely of odd permutations. However, every element of A_5 is conjugate to its inverse ([**12**, p. 60]). Examples of infinite groups with this property are given in [**7**].

Note that the element $\sigma = (1,2)(3,4)$ of S_4 has order 2, so $C(\sigma) = B(\sigma)$. Also, $\tau = (2,3,4) \in C(\sigma)$ is of order 3 and $\tau\sigma = \sigma^{-1}\tau$.

(iii) An Infinite Group. Let $G = \langle a, k \rangle$ be the finitely generated group subject to the relations $kak^{-1} = a^{-1}$ and $k^4 = e$. Then $C(a) = \langle a, k^2 \rangle$, and if $x \in B(a)$ then for some $n \in \mathbb{Z}$ either $x = ka^n$ or $x = k^{-1}a^n$. Now

$$(ka^n)^2 = k^2(k^{-1}a^nk)a^n = k^2a^{-n}a^n = k^2 \neq e,$$

and similarly for $k^{-1}a^n$. It follows that there are no involutions in $B(a)$. However, $|ka^n| = 4$ and $\{x^2 : x \in B(a)\}$ is a singleton set. It follows that $B(a)$ conjugates $C(a)$ to $C(a)^{-1}$.

(iv) Group Rotations. Let G be a compact *monothetic* topological group: there exists some $a \in G$ for which the set $\{a^n : n \in \mathbb{Z}\}$ is dense in G. In this case G is abelian, and we assume that $a^2 \neq e$. A nice example is the unit circle S^1 in the complex plane, with $a \in S^1$ chosen so that it is not a root of unity.

Let $\mathcal{G}$ be the group of all homeomorphisms $h : G \to G$, a subgroup of the permutation group of G. If we define $\phi_a : G \to G$ by $\phi_a(g) = a \cdot g$ for some fixed $a \in G$, then ϕ_a is a *rotation* of G and $\phi_a \in \mathcal{G}$.

We claim that $C(\phi_a) = \{\phi_b : b \in G\}$, for if $\psi \in C(\phi_a)$, then

$$\psi \circ \phi_a = \phi_a \circ \psi \Rightarrow \psi(ag) = a\psi(g), \quad \text{for all} \quad g \in G.$$

If $\psi(e) = b$, then $\psi(a) = a \cdot b$ and $\psi(a^n) = a^n \cdot b$ for all $n \in \mathbb{Z}$. Continuity of ψ now implies that $\psi(g) = b \cdot g$ for all $g \in G$, or $\psi = \phi_b$.

Note that $\phi_a^{-1}(g) = a^{-1}g$, so if we define $S: G \to G$ by $S(g) = g^{-1}$, then

$$S \circ \phi_a(g) = S(ag) = (ag)^{-1} = a^{-1}g^{-1} = \phi_a^{-1}(g^{-1}) = \phi_a^{-1} \circ S(g),$$

or $S \in B(\phi_a)$. It follows that every member of $B(\phi_a)$ is of the form $R = S \circ \phi_b$ for some $b \in G$. Now

$$R^2 = S \circ \phi_b \circ S \circ \phi_b = S \circ \phi_b \circ \phi_b^{-1} \circ S = S^2 = \mathrm{Id},$$

where Id is the identity in $\mathcal{G}$. We conclude that every member of $B(\phi_a)$ is an involution.

(v) Automorphism Groups. Let X be a group and let $\mathcal{G} = \mathrm{Aut}(X)$ be the automorphism group of X. If $\mathcal{G}$ is abelian (for example if X is cyclic, or is the p-adic integers) then $B(\phi) \neq \varnothing$ if and only if $\phi \in \mathcal{G}$ is an involution. In the general case there is a simple way to construct automorphisms conjugate to their inverses: Let $\phi \in \mathrm{Aut}(X)$. Then $\phi \times \phi^{-1} \in \mathrm{Aut}(X \times X)$. Define $\psi_0 \in \mathrm{Aut}(X \times X)$ by $\psi_0(x, y) = (y, x)$. Then $\psi_0 \in B(\phi \times \phi^{-1})$ and $\psi_0^2 = \mathrm{Id}$.

Other less-trivial members of $B(\phi \times \phi^{-1})$ can be described. For example, define $\psi \in \mathrm{Aut}(X \times X)$ by $\psi(x, y) = (y, \phi(x))$. We can verify directly that $(\phi \times \phi^{-1}) \circ \psi = \psi \circ (\phi^{-1} \times \phi)$, and ψ has infinite order if ϕ has infinite order.

Related examples can be given using countable direct products. Let X be a group and write $G = \Pi_{i=-\infty}^{\infty} X_i$, where $X_i = X$ for all $i \in \mathbb{Z}$. Define $\Phi: G \to G$ by $[\Phi(g)]_n = g_{n+1}$, where the subscript n denotes the nth coordinate. Then Φ is just the familiar shift automorphism.

Define automorphisms P and Q of G by $[P(g)]_n = g_{-n}$ and $[Q(g)]_n = g_{1-n}$. Then $\Phi = PQ$ is a product of two involutions. It follows from the last paragraph of Section 1 that Φ is conjugate to its inverse by an involution.

In a similar way we can construct other automorphisms of the product space that are conjugate to their inverses. Let $U: G \to G$ be defined by

$$U(\ldots, g_{-1}, \overset{*}{g}_0, g_1, g_2, \ldots) = \left(\ldots, \phi^{-1}(g_{-1}), \phi(g_0), \overset{*}{\phi}^{-1}(g_1), \phi(g_2), \ldots\right),$$

where the $*$ denotes the 0th coordinate and ϕ is an automorphism of X.

Note that $U^2 = \Phi^2$, where Φ is the shift map, so if we let $V = \Phi$ and $A = UV^{-1}$, that is,

$$A(\ldots, g_{-1}, \overset{*}{g}_0, g_1, \ldots) = \left(\ldots, \phi(g_{-1}), \overset{*}{\phi}^{-1}(g_0), \phi(g_1), \ldots\right),$$

then the same reasoning as in the preceding example implies that A is conjugate to A^{-1} and $U \in B(A)$.

5. DYNAMICAL ORIGINS. Let $T: X \to X$ be an *invertible measure-preserving transformation* on a Borel probability space $(X, \mathcal{F}, \mu)$. This means that $T^{-1}B \in \mathcal{F}$ and $\mu(T^{-1}B) = \mu(B)$ for all $B \in \mathcal{F}$. If T_1 and T_2 are two such transformations on corresponding probability spaces, T_1 is *conjugate* to T_2 if there is an invertible measure-preserving transformation $S: X_1 \to X_2$ such that $ST_1 = T_2S$ (when we write $f = g$ in this section, we mean $f(x) = g(x)$ for all $x \in X$ except possibly for a set of μ-measure zero). With every measure-preserving transformation T we associate a unitary operator

$$U_T: L^2(X, \mu) \to L^2(X, \mu); \quad U_T f(x) = f(Tx).$$

A measure-preserving transformation T is said to have *discrete spectrum* if U_T has discrete spectrum, i.e., there is a complete orthonormal sequence

$\{f_n\} \subset L^2(X, \mu)$ and a sequence $\{\lambda_n\}$ of complex numbers (of absolute value one) such that $f_n(Tx) = \lambda_n f_n(x)$ for all $n = 1, 2, \ldots$. More generally, T is said to have *simple spectrum* if there exists some $h \in L^2(X, \mu)$ such that the closed linear span of $\{U_T^n h : n \in \mathbb{Z}\}$ is all of $L^2(X, \mu)$.

A transformation T is *ergodic* if for any measurable function f, the condition $f(Tx) = f(x)$ implies $f =$ constant. This is equivalent to the condition: if $A \in \mathcal{F}$, then $T^{-1}A = A$ implies $\mu(A) = 0$ or 1.

Suppose G is a compact topological group with Haar measure λ, and whose measurable sets are the Borel subsets of G. If $a \in G$, the *rotation* $\phi_a : G \to G$ given by $\phi_a(g) = a \cdot g$ is a Haar measure-preserving transformation. The transformation ϕ_a is ergodic if and only if $\langle a \rangle$ is dense in G. A celebrated result of Halmos and von Neumann [**6**, Theorem 4] says that an ergodic transformation T with discrete spectrum is conjugate to an ergodic rotation $\phi_a : G \to G$ for some compact topological group G and some $a \in G$.

The characters of G have the property

$$\chi(\phi_a(g)) = \chi(a \cdot g) = \chi(a)\chi(g),$$

i.e., they are eigenfunctions of ϕ_a. Conversely, every eigenfunction of ϕ_a is a character of G. It is a consequence of Pontryagin duality theory that the characters constitute a complete orthonormal basis for $L^2(G, \lambda)$.

Let T be an ergodic transformation with discrete spectrum. The discrete spectrum theorem of Halmos and von Neumann implies that T is conjugate to its inverse. Also, if S is measure-preserving and $ST = T^{-1}S$, then $S^2 = I$. Furthermore, the centralizer of T (with a suitable topology) is a compact abelian group that is isomorphic to G in a natural way. Conversely, any ergodic T for which $C(T)$ is compact must have discrete spectrum.

Halmos and von Neumann asked whether every ergodic transformation is conjugate to its inverse via an involution. In 1951, Anzai showed that there is an ergodic transformation that is not conjugate to its inverse. As a consequence, this type of inverse-conjugacy problem lay dormant for some years. Recently it was shown that if a transformation T has simple spectrum and is conjugate to its inverse, then every conjugation is an involution. This generalizes aspects of the Halmos–von Neumann theorem [**3**, Theorem 1]. In addition, in [**3**], [**4**], [**8**], and [**9**] one finds many examples of transformations whose centralizers and skew centralizers have group theoretic properties similar to those discussed in this paper. Ergodic transformations with a wide variety of centralizers are now known. For an ergodic T, $C(T)$ can be a compact abelian group, a countable cyclic group, an uncountable monothetic group (hence abelian, but not locally compact), as well as many other abelian and non-abelian groups. A nice treatment of certain examples arising in this theory is given in [**2**].

6. CONCLUDING REMARKS. In the dynamical systems literature there has been some confusion about the role of involutory time-reversal systems. Some authors seem to believe that the set of reversing symmetries $B(a)$ (if non empty) must always contain an involution. Our examples show that this is not generally the case, although it is true in certain special contexts. For example, if $G = GL(n, \mathbf{R})$ then for all $a \in G$ such that $B(a) \neq \emptyset$, $B(a)$ contains an involution; see [**13**, Chapter 2]. Related results are given in [**5**], where properties of real orthogonal matrices are studied in this context. A survey of time-reversal symmetry in dynamical systems is given in [**11**].

ACKNOWLEDGMENTS. I thank Jay Zimmerman of Towson University and Jeroen Lamb of Warwick University for carefully reading the manuscript, suggesting many improvements, and providing additional references.

REFERENCES

1. H. S. M. Coxeter and W. O. J. Moser, *Generators and Relations for Discrete Groups*. Springer-Verlag, Berlin Heidelberg New York, 1984.
2. N. A. Friedman, Replication and stacking in ergodic theory, *Amer. Math. Monthly* **99** (1992) 31–41.
3. G. R. Goodson, A. del Junco, M. Lemańczyk, and D. J. Rudolph, Ergodic transformations conjugate to their inverses by involutions, *Ergodic Theory Dynamical Systems* **16** (1996) 97–124.
4. G. R. Goodson and M. Lemańczyk, Transformations conjugate to their inverses have even essential values, *Proc. Amer. Math. Soc.* **124** (1996) 2703–2710.
5. G. R. Goodson, The inverse-similarity problem for real orthogonal matrices, *Amer. Math. Monthly* **104** (1997) 223–230.
6. P. R. Halmos and J. von Neumann, Operator methods in classical mechanics II, *Annals Math.* **43** (1941) 332–350.
7. P. Hall, Some constructions for locally finite groups, *J. London Math. Soc.* **34** (1959) 305–319.
8. J. S. W. Lamb, Reversing symmetries in dynamical systems, *J. Phys. A: Math. Gen.* **25** (1992) 925–937.
9. J. S. W. Lamb, Reversing symmetries in dynamical systems, Ph.D. Thesis, University of Amsterdam, 1994.
10. J. S. W. Lamb and G. R. W. Quispel, Reversing k-symmetries in dynamical systems, *Phys. Rev. D* **73** (1994) 277–304.
11. J. S. W. Lamb and J. A. G. Roberts, Time-reversal in dynamical systems: a survey, University of Warwick preprint 34/1997.
12. W. Lederman, *Introduction to Group Characters*, Cambridge University Press, Cambridge, 1976.
13. M. B. Sevryuk, *Reversible systems*. Lecture Notes in Math. 1211, Springer-Verlag, Berlin Heidelberg New York, 1986.

GEOFFREY GOODSON was born in Cape Town, South Africa and was educated in Great Britain. He now teaches at Towson University (previously Towson State University) in Maryland. His interests include ergodic theory and mathematical exposition at the undergraduate level.
Towson University, Towson, MD 21252
ggoodson@towson.edu

More on Paperfolding

Dmitry Fuchs and Serge Tabachnikov

It is a common knowledge that folding a sheet of paper yields a straight line. We start our discussion of paperfolding with a mathematical explanation of this phenomenon. The model for a paper sheet is a piece of the plane; folding is an isometry of the part of the plane on one side of the fold to another, the fold being the curve of fixed points of this isometry (see Figure 1). The statement is that this curve is straight, that is, has zero curvature.

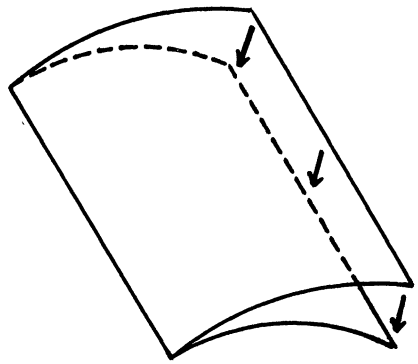


Figure 1

If not, consider an arc γ of the fold with nonvanishing curvature. Let γ_+ be the curve at (small) distance ε from γ on the concave side, and let γ_- be the corresponding curve on the convex side, as illustrated in Figure 2. Then

$$\textit{length}\,\gamma_+ > \textit{length}\,\gamma > \textit{length}\,\gamma_-,$$

where the difference between $\textit{length}\,\gamma_+$ and $\textit{length}\,\gamma_-$ is of order $\varepsilon \cdot \textit{length}\,\gamma \cdot \textit{curvature}\,\gamma$. On the other hand, the isometry takes γ_+ to γ_-, so $\textit{length}\,\gamma_+ = \textit{length}\,\gamma_-$. This is a contradiction.

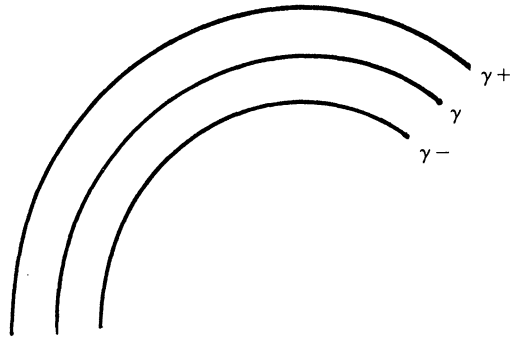


Figure 2

In spite of what has just been said, *one can fold paper along an arbitrary smooth curve*! The reader is invited to perform an experiment: draw a curve on a sheet of paper and slightly fold the paper along the curve. A word of practical advice: press hard when drawing the curve. It also helps to cut a neighborhood of the curve, for it is inconvenient to work with too large a sheet. A more serious reason for restricting to a neighborhood is that this way one avoids self-intersections of the sheets, unavoidable otherwise. The result looks somewhat like Figure 3(a):

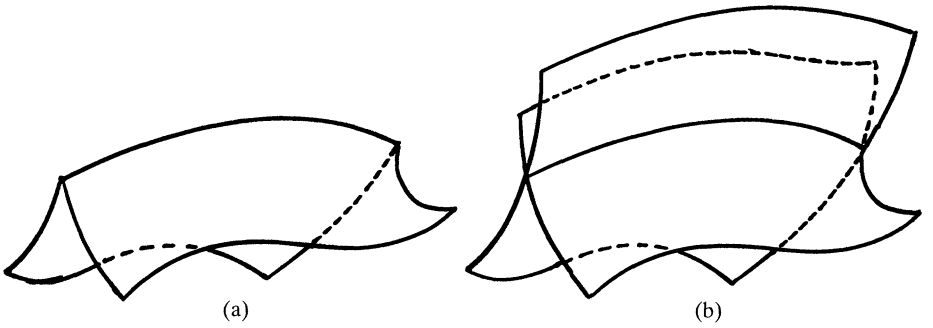


Figure 3

One may even start with a closed curve drawn on paper. To be able to fold, one has to cut a hole inside the curve; the result is shown in Figure 4.

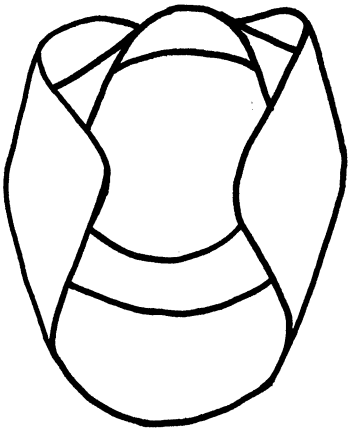


Figure 4

It goes without saying that the argument in the opening paragraphs of this article does not contradict the possibility of folding along a curve: the two sheets in Figure 3(a) meet at a nonzero angle. To fix terminology call the curve drawn on paper the *fold* and denote it by Γ; call the curve in space obtained by folding along Γ the *ridge* and denote it by γ. The above described experiments and dozens of similar ones that kept us busy lately reveal the following list of observations:

(1) *It is possible to start with an arbitrary smooth fold and obtain an arbitrary ridge provided the ridge is "more curved" than the fold.*

(2) *If the ridge is only slightly more curved than the fold, then the neighborhood of the fold to be folded should be taken very thin, at least from the side of the convex domain of the plane bounded by the fold.*
(3) *If the fold has an inflection point (i.e., point of zero curvature) then the corresponding point of the ridge is also an inflection point (notice that, unlike plane curves, a generic space curve does not have inflection points at all).*
(4) *If the fold is a closed strictly convex curve then the ridge has a nonzero torsion, that is, does not lie in one plane.*
(5) *If the fold is a nonclosed arc, the folded paper tends to occupy such a position that the ridge lies in a plane, and the angle made by the two sheets is constant along the ridge.*

What follows is an attempt to explain these experimental observations. A surface obtained by bending, without folding, a paper sheet is a *developable* surface, that is, a surface locally isometric to the plane (one cannot stretch paper!). The theory of such surfaces is due to Euler; its main results are as follows. A developable surface is a *ruled* surface, i.e., it consists of a one-parameter family of straight lines called *rulings*. These lines are not arbitrary: they are tangent to a certain space curve called the *edge of regression* (this description does not include two special cases, cylinders and cones, which are also developable surfaces). The tangent planes to a developable surface along every ruling coincide: one can put not only a knitting needle on such a surface but also a ruler. Thus a developable surface is the envelope of a one-parameter family of planes (see Figure 5).

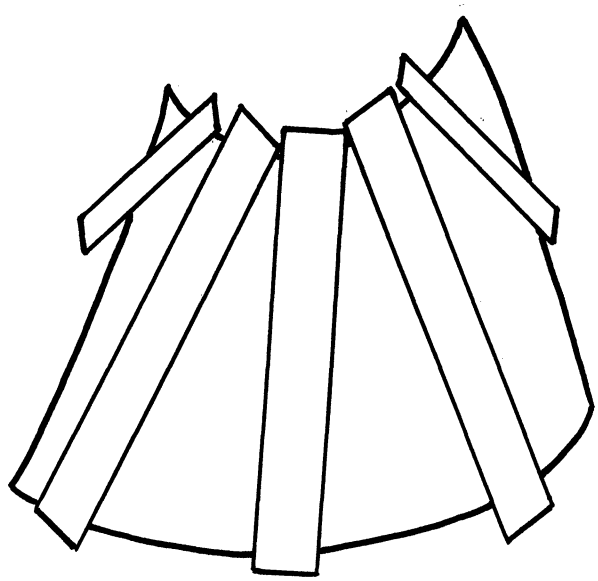


Figure 5

Consider Figure 3(b). One sees two developable surfaces intersecting along a space curve γ. Unfolding either of the surfaces to the plane transforms γ to the same plane curve Γ. Reverse the situation and pose the following question: given a plane curve Γ, a space curve γ and an isometry $f : \Gamma \to \gamma$, is it possible to extend f

to an isometric embedding of a plane neighborhood of Γ to space? Said differently, can one bend a sheet of paper with a curve Γ drawn on it so that Γ bends to a given space curve γ?

Theorem. *Assume that for every $x \in \Gamma$ the absolute value of the curvature of γ at point $f(x)$ is greater than that of Γ at x. Then there exist exactly two extensions of f to isometric embeddings of a plane neighborhood of Γ to space.*

These two embedded surfaces are the sheets intersecting along the ridge in Fig. 3(b). Extending the sheets beyond the ridge one obtains another configuration of sheets that meet along γ. Thus there are exactly two ways to fold paper along Γ to produce the curve γ. This explains and extends the first of the previously mentioned observations.

If γ lies in a plane, one of the sheets is obtained from another by reflection in this plane. In the general case of a nonplanar curve γ the tangent planes of the two sheets are symmetric with respect to the osculating plane of γ at its every point.

Proof: Parametrize the curves γ and Γ by the arclength parameter t so that $\gamma(t) = f(\Gamma(t))$. Let the desired developable surface S make the angle $\alpha(t)$ with the osculating plane of the curve $\gamma(t)$ (well defined since, by assumption, the curvature of γ never vanishes). Denote by $k(t)$ the curvature of the space curve γ and by $K(t)$ that of the plane curve Γ. The geodesic curvature vector of γ in S is the projection of the curvature vector of γ in space onto S; thus the geodesic curvature of γ equals $k(t)\cos\alpha(t)$. Since an isometry preserves the geodesic curvature of curves, $k\cos\alpha = K$. This equation uniquely determines the nonvanishing function $\alpha(t)$ up to the substitution $\alpha \to \pi - \alpha$. To construct the developable surface S from the function $\alpha(t)$, consider the plane through point $\gamma(t)$ that makes the angle $\alpha(t)$ with the osculating plane of γ. Such planes constitute a one-parameter family, and according to the above described general theory, their envelope is a developable surface. ■

Remarks. 1. The theorem is hardly new: it is mentioned as an exercise in [2] with a reference to [1]. For a later discussion of paperfolding see [3].

2. A direct computation involving the Frenet formulas for γ (which we omit) makes it possible to find the angle $\beta(t)$ made by the rulings $l(t)$ with the curve $\gamma(t)$ in terms of the torsion $\kappa(t)$ of γ:

$$\cot\beta(t) = \frac{\alpha'(t) - \kappa(t)}{k(t)\sin\alpha(t)}. \tag{1}$$

For the two developable surfaces corresponding to the angles $\alpha(t)$ and $\pi - \alpha(t)$ one has:

$$\cot\beta_1(t) + \cot\beta_2(t) = -\frac{2\kappa(t)}{k(t)\sin\alpha(t)}.$$

Therefore the ridge γ is a plane curve (i.e., $\kappa = 0$) if and only if $\beta_1 + \beta_2 = \pi$. In this case, unfolding the two sheets on the plane yields the straight rulings that extend each other on both sides of the fold Γ; see Figure 6.

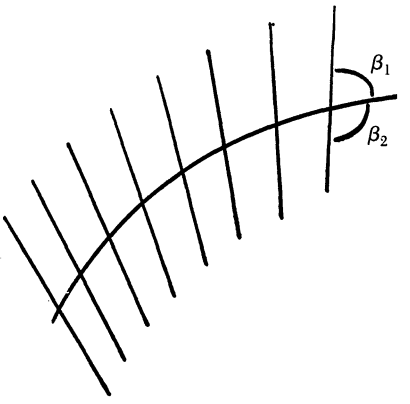


Figure 6

The reader with a taste for further experimentation may find the following one of interest. Start with a fold Γ and tape a number of pins on both its sides. In this way one prescribes the angles $\beta_1(t)$ and $\beta_2(t)$. Then fold along Γ, as illustrated in Figure 7.

Remark 2 may be used to explain the second of our experimental observations. Namely, since $\cos\alpha(t) = K(t)/k(t)$, the formula (1) for $\cot\beta(t)$ may be rewritten as

$$\cot\beta(t) = \frac{\alpha'(t) - \kappa(t)}{\sqrt{k(t)^2 - K(t)^2}}.$$

Assume that $K(t)$ is close to $k(t)$; then $\alpha(t)$ is close to zero. We need, however, to assume that $\alpha'(t)$ is also close to zero. If, moreover, $\kappa(t)$ is bounded away from zero, then $\alpha'(t)$ is small, and $\cot\beta(t)$ is large; hence the angle $\beta(t)$ is small. It is clear that straight lines crossing the boundary of a convex domain in the plane at small angles cannot penetrate deep in the domain; hence, they have to cross each other near the boundary (see Figure 8).

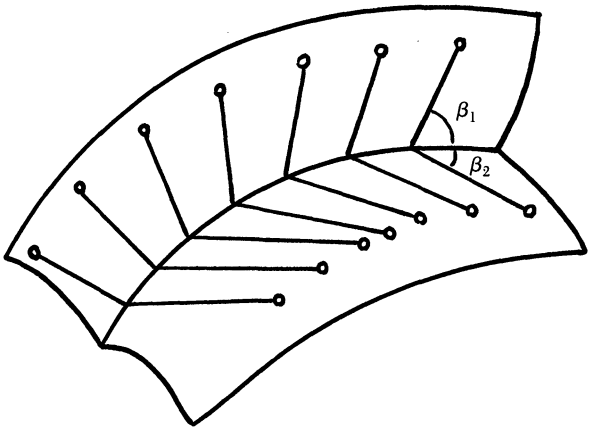


Figure 7

But the rulings of a non-self-intersecting developable surface do not cross each other. Hence, to avoid crossings we need to make the neighborhood of the fold

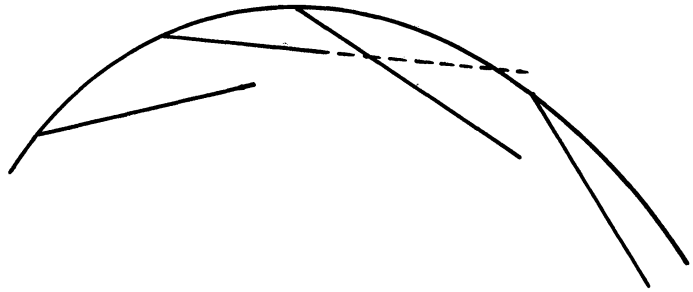


Figure 8

thin. The limit case of this observation is particularly interesting. Suppose that $K = k$. Then the isometry between the fold and the ridge cannot be extended into the convex domain bounded by the fold at all. It can be extended into the concave domain, and we get a developable surface, for which *the edge is the edge of regression*. Indeed, formula (1) for $\cot\beta(t)$ gives $\cot\beta(t) = \infty$; hence $\beta(t) = 0$, and the rulings of the surface are all tangent to the ridge. Of course, in this way we get only one of the two pieces of the surface, cut along the edge of regression. The other piece may be made of another copy of the same concave domain. The difference between the two pieces is that for each tangent to the boundary of our concave domain, divided into two halves by the point of tangency, one half is straight on one of the pieces and the other half is straight on the other piece. The image of the whole tangent on each piece is a curve, half of which is straight and half of which is curved, as illustrated in Figure 9.

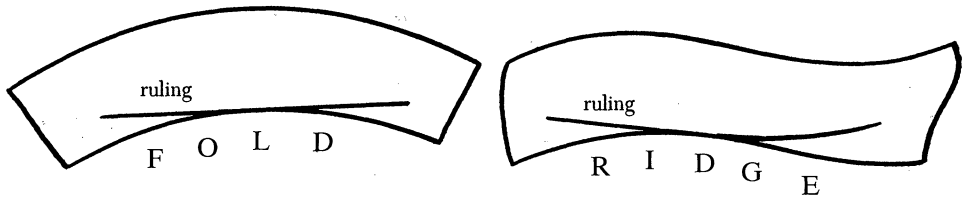


Figure 9

The union of the two pieces with the images of the two tangents looks like Figure 10.

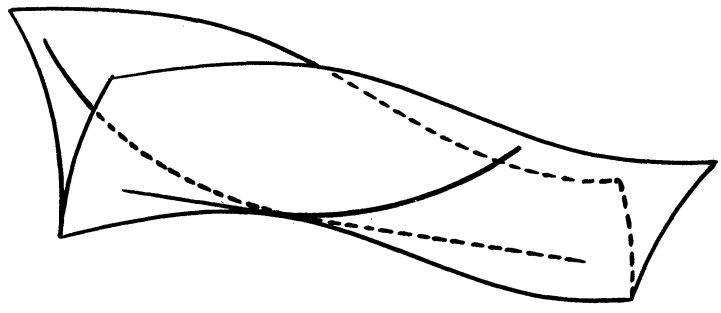


Figure 10

As a by-product of these observations we learn how to make a paper model of a developable surface with a prescribed edge of regression (without inflection points). To do this we should draw a planar curve whose curvature is precisely the same as that of the intended edge of regression; two copies of this drawing should be made on two separate sheets of paper. Then we cut the sheets along the curves and take the concave portions of both. After this we bend the two pieces to make their edges fit into the given spatial curve. This may be done in two different ways, and we must bend our (identical) pieces into different surfaces; these two surfaces comprise the developable surface we are constructing. Since the "angle" between the two pieces should be 0, it may be useful to glue the two pieces before bending along a very thin neighborhood of the edges. But be aware, that this bending is not even a twice differentiable mapping (this is why we used quotation marks for the word "angle"), and the paper will be resistant to this construction. Pins, attached to the two pieces tangentially to the edges (as shown on Figure 11) may help.

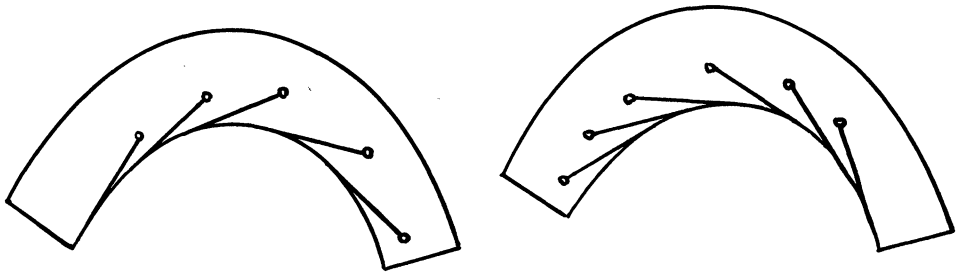


Figure 11

Back to our list of experimental observations. The first two have been explained, proceed to the third one. Let $\Gamma(t_0)$ be a nondegenerate inflection point, so the fold looks like a cubic parabola near this point (see Figure 12).

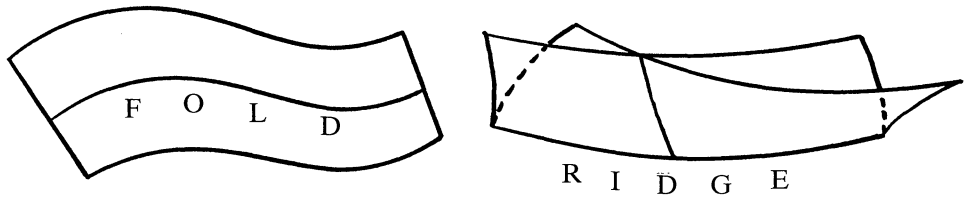


Figure 12

Then $K(t_0) = 0$ and, according to the already familiar formula $k\cos\alpha = K$, either $\alpha(t_0) = \pi/2$ or $k(t_0) = 0$. We want to show that the latter possibility holds. Suppose not; then both sheets are perpendicular to the osculating plane of γ at

point $\gamma(t_0)$ and, therefore, coincide. Moreover, if $k(t_0) \neq 0$ then the projection of the curvature vector of the space curve γ onto each sheet is the vector of the geodesic curvature therein. This vector lies on one side of γ on the surface at points of the form $\gamma(t_0 - \varepsilon)$ just before the inflection point and on the other side at points $\gamma(t_0 + \varepsilon)$ just after it. Therefore the function $\alpha(t) - \pi/2$ changes sign at $t = t_0$. This means that the two sheets pass through each other at $t = t_0$. Possible in the class of immersions, this cannot happen with real paper. Thus $k(t_0) = 0$, that is, the ridge has an inflection point.

Next, consider the fourth observation of our list. Assume that both γ and Γ are closed plane curves and Γ is strictly convex. The relation between the curvatures still holds: $k\cos\alpha = K$, and $K(t)$ does not vanish. Hence $k(t) \geq K(t)$ for all t and $\int k(t) > \int K(t)$ unless $\alpha(t)$ identically vanishes. On the other hand, the integral curvature of a simple closed plane curve equals 2π, so the above integrals must be equal. This is a contradiction. It is interesting that if Γ is closed *nonconvex* curve, one can nontrivially bend paper along Γ keeping Γ in the plane; see Figure 13.

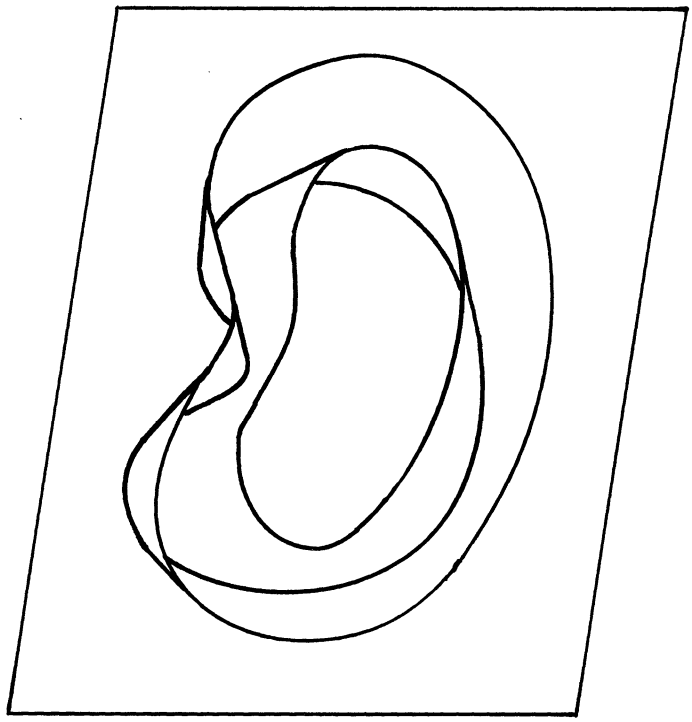


Figure 13

Finally we turn to the fifth experimantal observation. This one takes us into dangerous waters because its explanation requires further assumptions concerning elasticity properties of paper. A strip of paper resists twisting: being relaxed it tends to become flat. Consider a space curve $\gamma(t)$ parametrized by the arclength. Let S be a thin strip along γ and $v(t)$ the unit normal vector field to γ in S. Define the *twist* of S to be the length of the projection of the vector $v'(t)$ to the normal plane of $\gamma(t)$. Our assumption is that a paper strip tends to minimize its twist. Let the strip make the angle $\alpha(t)$ with the osculating plane of the curve $\gamma(t)$.

Then a computation, similar to the one mentioned in Remark 2, gives the value $|\kappa(t) - \alpha'(t)|$ for the twist. Folding paper, one produces two strips along the ridge $\gamma(t)$, the angles being $\alpha(t)$ and $\pi - \alpha(t)$. The twists of these strips are equal to $|\kappa - \alpha'|$ and $|\kappa + \alpha'|$. Both quantities attain minimum if $\kappa(t) = 0$ and $\alpha(t)$ is constant. This appears to explain the fifth experimental observation.

ACKNOWLEDGMENTS. We are grateful to M. Kontsevich who taught us how to fold paper along curves and brought the problem to our attention, and to R. Montgomery for an explanation of mathematical models of elasticity. The second author was supported in part by NSF; the paper was written while he enjoyed the hospitality of the MSRI and MPIM.

REFERENCES

1. L. Bianchi, *Vorlesungen über Differentialgeometrie*, Leipzig, B. G.Teubner, 1899.
2. W. Blaschke, *Vorlesungen und Geometrische Grundlagen Einsteins Relativitäts-theorie*, Berlin, J. Springer, 1930.
3. D. A. Huffman, Curvature and Creases: A Primer on Paper. IEEE Transactions on Computers, **C-25**, No. 10 (1976) 1010–1019.

DMITRY FUCHS was at Moscow State University from 1955 to 1990, first as a student (Ph.D., 1964), then as a professor. Since 1991 he has been a professor at the University of California, Davis. He works in topology, representation theory, homological algebra, and symplectic topology.
University of California, Davis CA 95616
fuchs@comp.ucdavis.edu

SERGE TABACHNIKOV received his Ph.D. from Moscow State University in 1987; he was a student of D. Fuchs. In 1988–90 he was the head of the Mathematical Department of *Kvant* magazine for which he wrote about 20 expository papers (*Kvant*'s English-language twin, *Quantum* is published by Springer-Verlag). He has been at the University of Arkansas, Fayetteville since 1990, with occasional leaves to visit research institutes in Europe. His interests include symplectic geometry, dynamical systems, and knot theory.
University of Arkansas, Fayetteville AR 72701
serge@math.uark.edu

Cicero on "pure" vs. "applied" science . . .

Moreover, Catulus, if you ask me my personal opinion as to the study in question, I do not think that a person of ability, and acquainted at first hand with public life and procedure in parliament and the law-courts, requires as much time as has been taken for themselves by those who have spent the whole term of their life in study. For all branches of knowledge are handled by those who apply them to practice in a different manner from that in which they are handled by those who take their pleasure in the pursuit of the sciences themselves and have no intention of following any other career.

De Oratore, III. xxiii. 86
Contributed by Adi Ben-Israel, Rutgers University

Trigonometric Integrals and Hadamard Products

L. R. Bragg

1. INTRODUCTION. Let $f(z) = \sum_{n=0}^{\infty} a_n z^n$ and $g(z) = \sum_{n=0}^{\infty} b_n z^n$ be analytic functions in respective disks D_1 and D_2 centered at the origin. Let $h(\xi) = \sum_{n=0}^{\infty} a_n b_n \xi^n$ be analytic in some disk centered at $\xi = 0$. In a famous 1899 paper [**5**], J. Hadamard used the integral convolution

$$h(\xi) = (2\pi)^{-1} \int_0^{2\pi} f(ze^{i\theta}) g(\zeta e^{-i\theta})\, d\theta \tag{1.1}$$

($\xi = z\zeta$) to characterize the singularities of the function $h(\xi)$ in terms of those of $f(z)$ and $g(z)$ (see also [**13**, p. 157]). The coefficient-wise product of power series denoted by $h(\xi) = f(z) \circ g(\zeta)$ is often called the *Hadamard product* of $f(z)$ and $g(\zeta)$ (alternative appellations include *Schur product* and *quasi inner product*).

The Hadamard product appears naturally in a variety of theoretical and applied mathematical questions. While studying the Bieberbach conjecture long before L. deBrange's proof, C. Loewner and E. Netanyahu [**7**] showed that a Hadamard-type product of two normalized ($f(0) = 0$, $f'(0) = 1$) univalent (one-to-one) analytic functions in the unit disk need not be univalent and can even have a zero derivative. St. Ruschewegh and T. Sheil-Small [**12**] showed that the Hadamard product of two normalized analytic convex mappings of the unit disk is a convex mapping. In [**3**], the Hadamard product appears in a variety of multiplier problems. In an 1894 paper [**9**], Th. Moutard used (essentially) the fact that an entry-wise product of positive definite matrices is positive definite to establish uniqueness theorems for solutions of a class of elliptic partial differential equations. Many properties and applications of the Hadamard product are surveyed in [**6**]. The author has used this product and a generalization to derive known and new formulas for special functions [**1**]. Finally, the Hadamard product, together with a factor switching property, was used in [**2**] to construct solution formulas for a variety of Cauchy-type problems.

Because of its connection with the integral convolution (1.1), the Hadamard product can lead to elegant evaluations of complicated trigonometric integrals and provide analytic derivations of combinatorial identities. The student with a solid background in the calculus and basic differential equations can use it to carry out beginning studies on special functions and their representations while being introduced to notions of complex variables. We focus on these areas while emphasizing how to choose the functions in the associated Hadamard products.

For example, consider the series definition for the Bessel function

$$J_0(z) = \sum_{n=0}^{\infty} (-1)^n z^{2n} / \{2^{2n} n! \cdot n!\} = \sum_{n=0}^{\infty} (1/(2^n n!))\big((-1)^n/(2^n n!)\big) z^{2n}$$

and observe that the terms $1/(2^n n!)$ are the coefficients in the series for $e^{z/2}$ while the terms $(-1)^n/(2^n n!)$ are the coefficients in the series for $e^{-z/2}$. It follows

that

$$J_0(z) = e^{z/2} \circ e^{-z/2} = (2\pi)^{-1} \int_0^{2\pi} e^{(ze^{i\theta})/2} e^{-(ze^{i\theta})/2}\, d\theta.$$

Applying the Euler formulas to the integrand yields the standard integral formula

$$J_0(z) = (2\pi)^{-1} \int_0^{2\pi} \cos(z\sin\theta)\, d\theta.$$

Before outlining the topics to be considered in the following sections, we introduce a Hadamard-type product depending on two integer parameters.

Definition 1. *For p and q relatively prime, and for $z \in D_1$, $\zeta \in D_2$, define*

$$f(z)_p \circ_q g(\zeta) = (2\pi)^{-1} \int_0^{2\pi} f(ze^{pi\theta}) g(\zeta e^{-qi\theta})\, d\theta = \sum_{n=0}^{\infty} a_{qn} b_{pn} z^{qn} \zeta^{pn}. \quad (1.2)$$

If either $f(z)$ or $g(\zeta)$ is a polynomial, then the series in (1.1) *is a polynomial in $z\zeta$.*

In Section 2, *we apply* (1.1) *in various ways to polynomials of the form* $(1 \pm z)^n$ *to obtain evaluations of numerous trigonometric integrals and prove binomial identities. Repeated Hadamard products are employed in Section* 3 *to evaluate multiple trigonometric integrals. Section* 4 *treats integral representations of some special functions in which at least one of the functions entering the associated Hadamard product is not a polynomial. Finally, in Section* 5, *we introduce the selector function, which sums the coefficients of a subseries of a given series, and show how it can be used to replace certain sums by integrals.*

2. SPECIAL SUMS AND SINGLE INTEGRALS. *Let us now use* (1.1) *to evaluate some special sums and trigonometric integrals.*

Example 2.1. *We wish to establish the familiar binomial identity*

$$\sum_{j=0}^{2n} (-1)^j \binom{2n}{j}^2 = (-1)^n \binom{2n}{n} \quad (2.1)$$

(*see* **[11]** *for this and related identities*). *On the one hand, through integration by parts, one obtains a reduction relation which permits showing*

$$\int_0^{2\pi} \sin^{2n}\theta\, d\theta = \frac{2\pi}{2^{2n}} \binom{2n}{n}. \quad (2.2)$$

Alternatively, we note that $2i\sin\theta = e^{i\theta} - e^{-i\theta} = (1 + e^{i\theta})(1 - e^{-i\theta})$, *which suggests considering the Hadamard product* $(1 + z)^{2n} \circ (1 - z)^{2n}$. *With* $a_j = \binom{2n}{j}$, $b_j = (-1)^j a_j$, *and* $p = q = 1$, (1.1) *gives*

$$(1 + z)^{2n} \circ (1 - z)^{2n} = \frac{1}{2\pi} \int_0^{2\pi} (1 + ze^{i\theta})^{2n} (1 - ze^{-i\theta})^{2n} d\theta$$

$$= \sum_{j=0}^{2n} (-1)^j \binom{2n}{j}^2 z^{2j}. \quad (2.3)$$

Setting $z = 1$, we have

$$\int_0^{2\pi} \sin^{2n}\theta \, d\theta = (-1)^n \frac{2\pi}{2^{2n}} \sum_{j=0}^{2n} (-1)^j \binom{2n}{j}^2, \tag{2.4}$$

which, together with (2.2) *gives* (2.1).

Example 2.2. *Suppose that p and q are relatively prime and let r be a positive integer. We prove*

$$\int_0^{\pi} \cos^{pr}(q\theta) \cdot \cos^{qr}(p\theta) \, d\theta = \frac{\pi}{2^{r(p+q)}} \sum_{l=0}^{r} \binom{qr}{ql}\binom{pr}{pl}. \tag{2.5}$$

Replace $\cos\theta$ by $(e^{i\theta} + e^{-i\theta})/2$ and take $f(z) = (1+z)^{qr}$ and $g(z) = (1+z)^{pr}$ in (1.2), *which becomes*

$$f(z)_p \circ_q g(z) = (2\pi)^{-1} \int_0^{2\pi} (1 + ze^{pi\theta})^{qr} (1 + ze^{-qi\theta})^{pr} \, d\theta$$

$$= \sum_{l=0}^{r} \binom{qr}{ql}\binom{pr}{pl} z^{l(p+q)}. \tag{2.6}$$

Compute

$$(1 + e^{pi\theta})^{qr}(1 + e^{-qi\theta})^{pr}$$

$$= \left[e^{pqri\theta/2}(e^{pi\theta/2} + e^{-pi\theta/2})^{qr}\right]\left[e^{-qpri\theta/2}(e^{qi\theta/2} + e^{-qi\theta/2})^{pr}\right]$$

$$= \left(2\cos\frac{p\theta}{2}\right)^{qr}\left(2\cos\frac{q\theta}{2}\right)^{pr} = 2^{r(p+q)}\cos^{qr}\left(\frac{p\theta}{2}\right)\cos^{pr}\left(\frac{q\theta}{2}\right),$$

insert this into the integral in (2.6) (*with $z = 1$*), *make the change of variables $\theta = 2\phi$, and* (2.5) *follows.*

A generalization of (2.5) *can be obtained by taking $f(z) = (1+z)^m$ and $g(z) = (1+z)^n$. By applying* (1.1), *taking the real part of the integral involved, and making a simple change of variables, one can establish*

$$\int_0^{\pi} (\cos p\phi)^m (\cos q\phi)^n \cos(mp - nq)\phi \, d\phi = \frac{\pi}{2^{m+n}} \sum_{l=0}^{H} \binom{m}{ql}\binom{n}{pl} \tag{2.7}$$

where $H = \min\left\{\left[\frac{m}{q}\right], \left[\frac{n}{p}\right]\right\}$. A variety of analogous integrals appear in [**4**, *pp.* 372–374].

Example 2.3. *Let $z_1 = z_2 = z$ and take $f(z) = g(z) = (a + bz + az^2)^n$. The trinomial expansion gives $f(z) = \sum_{k=0}^{2n} c_k z^k$, in which*

$$c_k = \sum_{j_2 + 2j_3 = k} \binom{n}{j_1, j_2, j_3} a^{j_1 + j_3} b^{j_2}$$

where $j_1 + j_2 + j_3 = n$ and $j_i \geq 0$. We show that

$$\sum_{k=0}^{2n} c_k^2 = b^{2n}{}_2F_1(-n, -n + 1/2; 1; 4a^2/b^2). \tag{2.8}$$

where ${}_2F_1(\alpha, \beta; \gamma; z)$ denotes the usual Gauss hypergeometric function with parame-

ters α, β, *and* γ [**8**, *pp.* 37–62]. *Compute*

$$\sum_{k=0}^{2n} c_k^2 = (f(z) \circ f(z))|_{z=1}$$

$$= (2\pi)^{-1} \int_0^{2\pi} (a + be^{i\theta} + ae^{2i\theta})^n (a + be^{-i\theta} + ae^{-2i\theta})^n \, d\theta$$

$$= (2\pi)^{-1} \int_0^{2\pi} (2a\cos\theta + b)^{2n} \, d\theta. \tag{2.9}$$

Expand the integrand in the latter integral, integrate term-by-term, and use the beta function and its reduction properties (*see* [**8**, *pp.* 7–8]) *to find that*

$$\sum_{k=0}^{2n} c_k^2 = \sum_{k=0}^{n} \frac{(2n)! \cdot a^{2j} b^{2n-2j}}{(2n-2k)!k!k!}. \tag{2.10}$$

Finally, replace $(2n)!/(2n-2k)!$ *in this by* $2^{2k}(-n)_k(-n+1/2)_k$ *to obtain* (2.8).

3. SOME MULTIPLE TRIGONOMETRIC INTEGRALS. *The Hadamard product of two polynomials is a new polynomial, which one can use to form a Hadamard product with another polynomial, etc. By repeating the Hadamard product operation and finally evaluating at* $z = 1$, *as in Section* 2, *one can evaluate various multiple trigonometric integrals.*

Example 3.1. *As in* (2.1.6), *we have*

$$(1+z)^n \circ (1+z)^n = (2\pi)^{-1} \int_0^{2\pi} (1 + ze^{i\theta})^n (1 + ze^{-i\theta})^n \, d\theta$$

$$= \sum_{j=0\}}^{n} \binom{n}{j}^2 z^{2j} = F(z). \tag{3.1}$$

An application of (1.1) *to* $F(z) \circ (1+z)^n$ *yields*

$$\frac{1}{(2\pi)^2} \int_0^{2\pi} \int_0^{2\pi} \left[(1 + ze^{i(\phi+\theta)})(1 + ze^{i(\phi-\theta)})(1 + ze^{-i\phi})\right]^n d\phi \, d\theta$$

$$= \sum_{j=0}^{\left[\frac{n}{2}\right]} \binom{n}{j}^2 \binom{n}{2j} z^{4j}. \tag{3.2}$$

Set $z = 1$ *and observe that the bracketed term in the integrand can be written as* $2^3 e^{i\phi/2} \cos((\phi+\theta)/2) \cdot \cos((\phi-\theta)/2)\cos(\phi/2) = 2^2 e^{i\phi/2}\cos(\phi/2)[\cos\phi + \cos\theta]$. *Insert this into* (3.2), *with* $z = 1$, *and take the real part of both sides to get*

$$\int_0^{2\pi} \int_0^{2\pi} \cos(n\phi/2)\cos^n(\phi/2)[\cos\phi + \cos\theta]^n \, d\phi \, d\theta = \frac{4\pi^2}{2^{2n}} \sum_{j=0}^{\left[\frac{n}{2}\right]} \binom{n}{j}^2 \binom{n}{2j}. \tag{3.3}$$

A similar treatment of $G(z) = F(z) \circ F(z)$ *leads to the triple integral formula*

$$\int_0^{2\pi}\int_0^{2\pi}\int_0^{2\pi}[(\cos\theta_1 + \cos\theta_2)(\cos\theta_1 + \cos\theta_3)]^n d\theta_1\, d\theta_2\, d\theta_3 = \frac{(2\pi)^3}{2^{2n}} \sum_{k=0}^{n} \binom{n}{k}^4. \tag{3.4}$$

Example 3.2. *The repeated Hadamard product* $[(1+z)^m \circ (1+z)^n] \circ (1+z^2)^p$ *yields the double integral formula*

$$\int_0^{\pi}\int_0^{\pi} \cos[(m-n)\theta + (m+n-2p)\phi]\cos^m(\theta+\phi)\cos^n(\theta-\phi)\cos^p(2\phi)\, d\theta\, d\phi$$

$$= \frac{\pi^2}{2^{m+n+p}} \sum_{j=0}^{\min(m,n,p)} \binom{m}{j}\binom{n}{j}\binom{p}{j}. \tag{3.5}$$

Taking $m = n = p$, *it is not difficult to establish*

$$\int_0^{2\pi}\int_0^{2\pi}[\cos\theta(\cos\theta + \cos\phi)]^m d\theta\, d\phi = \frac{4\pi^2}{2^{2m}} \sum_{j=0}^{m} \binom{m}{j}^3.$$

4. SOME SPECIAL POLYNOMIALS. *Thus far, both Hadamard product factors* $f(z)$ *and* $g(\zeta)$ *have been polynomials. A convenient non-polynomial choice for* $f(z)$ *is the exponential function and we shall use it to obtain trigonometric integral representations for the classical Laguerre polynomials and for some special hypergeometric type polynomials.*

Example 4.1. *The Laguerre Polynomials. The Laguerre polynomials are defined by* $L_n(z) = \sum_{j=0}^{n} \frac{(-1)^j}{j!} \cdot \binom{n}{j} z^j$ (*see* [**8**, *p.* 239] *or* [**10**, *pp.* 200–217]). *We identify* $\binom{n}{j}$ *as the Maclaurin coefficients of* $(1+\zeta)^n$ *and* $(-1)^j/j!$ *as the Maclaurin coefficients of* e^{-z}. *Then using* (1.1), *we have*

$$L_n(z\zeta) = \sum_{j=0}^{n} \left\{\frac{(-1)^j z^j}{j!}\right\}\left\{\binom{n}{j}\zeta^j\right\} = e^{-z} \circ (1+\zeta)^n$$

$$= \frac{1}{2\pi}\int_0^{2\pi} e^{-ze^{i\theta}}(1 + \zeta e^{-i\theta})^n d\theta. \tag{4.1}$$

If we now select $z = x$ (*real*) *and* $\zeta = 1$, *we obtain*

$$L_n(x) = \mathrm{Re}\left\{(2\pi)^{-1}\int_0^{2\pi} e^{-x\cos\theta}\cdot e^{-ix\sin\theta - n\theta i/2}(e^{i\theta/2} + e^{-i\theta/2})^n\, d\theta\right\}$$

$$= 2^n(2\pi)^{-1}\int_0^{2\pi} e^{-x\cos\theta}\cos(x\sin\theta + n\theta/2)\cos^n(\theta/2)\, d\theta$$

$$= 2^n\pi^{-1}\int_0^{\pi} e^{-x\cos 2\phi}\cos(x\sin 2\phi + n\phi)\cos^n\phi\, d\phi. \tag{4.2}$$

Since $L_n(0) = 1$, *this yields* $\int_0^{\pi}\cos(n\phi)\cos^n\phi\, d\phi = \pi/2^n$.

Example 4.2. *A Hypergeometric Polynomial. Select* $f(z) = e^{-z}$ *and* $g(\zeta) = (1 - \zeta)^n$. *Then* (1.1) *gives*

$$e^{-z}{}_1 \circ {}_2(1 - \zeta)^n = (2\pi)^{-1} \int_0^{2\pi} e^{-ze^{2i\theta}} (1 - \zeta e^{-i\theta})^n \, d\theta$$

$$= \sum_{k=0}^{\left[\frac{n}{2}\right]} \frac{(-1)^k}{k!} \binom{n}{2k} z^k \zeta^k = p_n(z, \zeta). \tag{4.3}$$

Choose $z = x$, x *real, and* $\zeta = 1$ *and obtain*

$$p_n(x, 1) = \sum_{k=0}^{\left[\frac{n}{2}\right]} \frac{(-1)^k}{k!} \frac{n!}{(2k)!(n - 2k)!} x^k.$$

Taking the real and imaginary parts of the integral in (4.3), *and making the change of variables* $\theta = 2\phi$ *gives*

$$p_n(x, 1) = \begin{cases} \dfrac{(-1)^{n/2} 2^n}{\pi} \displaystyle\int_0^{\pi} e^{-x\cos 4\phi} \cos(x \sin 4\phi + n\phi) \sin^n \phi \, d\phi & \text{for } n \text{ even}, \\ \dfrac{(-1)^{(n-1)/2} 2^n}{\pi} \displaystyle\int_0^{\pi} e^{-x\cos 4\phi} \sin(x \sin 4\phi + n\phi) \sin^n \phi \, d\phi & \text{for } n \text{ odd}. \end{cases} \tag{4.4}$$

5. THE SELECTOR FUNCTION. *Let* $S_k(z) = \sum_{j=0}^{k} z^j = (1 - z^{k+1})/(1 - z)$ *and let* $P_n(z) = \sum_{j=0}^{n} a_j z^j$. *Form* $P_n(z)_1 \circ {}_q S_k(z)$ *and replace* z *by* 1 *to obtain*

$$(2\pi)^{-1} \int_0^{2\pi} P_n(e^{i\theta}) \frac{(1 - e^{-q(k+1)i\theta})}{(1 - e^{-qi\theta})} \, d\theta =$$

$$(2\pi)^{-1} \int_0^{2\pi} P_n(e^{i\theta}) e^{-qki\theta/2} \frac{\sin q(k + 1)\theta/2}{\sin q\theta/2} d\theta = \sum_{j=0}^{\min\left\{\left[\frac{n}{q}\right], \left[\frac{k}{q}\right]\right\}} a_{qj}. \tag{5.1}$$

The latter series is a sum of a subset of the coefficients of the polynomial $P_n(z)$ *in which the indices are multiples of* q. *This particular subset also depends upon the choice of* k. *It is clear that if* $q = 1$ *and* $k \geq n$, *then the series in* (5.1) *reduces to* $P(1)$. *Thus, with appropriate choices of* q *and* k, (5.1) *can be applied in a variety of ways to replace finite sums by trigonometric integrals. Because of its utility for singling out particular terms of a polynomial, we refer to* $S_k(z)$ *as a selector function.*

To write an integral formula for the binomial sum $\sum_{j=0}^{m} \binom{n}{j}$, take $P_n(z) = (1 + z)^n$, $q = 1$, and $k = m$ in (5.1). Applying the Euler relations and taking $\theta = 2\phi$ leads to

$$\sum_{j=0}^{m} \binom{n}{j} = \frac{2^n}{\pi} \int_0^{\pi} \cos^n(\phi) \cos((n - m)\phi) \frac{\sin((m + 1)\phi)}{\sin(\phi)} \, d\phi. \tag{5.2}$$

While the denominator in the integrand vanishes at $\phi = 0$ and $\phi = \pi$, so also does $\sin((m + 1)\phi)$. The singularities cancel and the integral (5.2) is well defined.

As a final example, take $P_n(z) = (1 + z)^{pq}$ and $k = p$ in (5.1) with $q \geq 1$. We leave it to the reader to conclude that

$$\int_0^{\pi} \cos^{pq}\phi \frac{\sin q(p+1)\phi}{\sin q\phi}\, d\phi = \frac{2\pi}{2^{pq}} \sum_{k=0}^{p} \binom{pq}{kq}. \tag{5.3}$$

REFERENCES

1. L. R. Bragg, Quasi inner products of analytic functions with applications to special functions, *SIAM J. Math. Anal.* **17** (1986) 220–230.
2. L. R. Bragg, A quasi inner product approach for constructing solution representations of Cauchy problems, *Rocky Mountain J. Math* **24** (1994) 1273–1306.
3. P. L. Duran, B. W. Romberg, and A. L. Shields, Linear functionals on H^p-spaces with $0 < p < 1$, *J. Reine Angew. Math.* **238** (1969) 32–60.
4. I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products*, Academic Press, Inc., New York, 1980.
5. J. Hadamard, Theoreme sur les series entieres, *Acta Math.* **22** (1899) 55–63.
6. R. A. Horn, The Hadamard product, *Proc. Symposia in Applied Math.* **40** (1990) 87–169.
7. C. Loewner and E. Netanyahu, On some compositions of Hadamard type in classes of analytic functions, *Bull. Amer. Math. Soc.* **65** (1959) 284–286.
8. W. Magnus, F. Oberhettinger, and R. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics*, Springer-Verlag New York, 1966.
9. Th. Moutard, Notes sur les Equations Derivees Partielles, *J. de L'Ecole Polytechnique* **64** (1894) 55–69.
10. E. Rainville, *Special functions*, MacMillan, New York, 1960.
11. J. Riordan, *Combinatorial Identities*, John Wiley and Sons, New York, 1968.
12. St. Ruscheweyh and T. Sheil-Small, Hadamard products of schlict functions and the Pólya-Schoenberg conjecture, *Comment. Math. Helv.* **48** (1973) 119–135.
13. E. C. Titchmarsh, *The Theory of Functions*. Oxford University Press, Oxford, 1949.

LOUIS R. BRAGG received his doctorate from the University of Wisconsin-Madison. He has reached the status of professor emeritus of mathematical sciences at Oakland University after having served there as a professor during 1966–1997. His primary research area is partial differential equations with an emphasis on transmutation and complex variable studies. He is also interested in special functions and parametric methods.

Department of Mathematical Sciences, Oakland University, Rochester, MI, 48309-4401.
bragg@oakland.edu

Cluster Primes

Richard Blecksmith, Paul Erdős, and J. L. Selfridge

1. INTRODUCTION. A prime $p > 2$ is called a *cluster prime* if every even positive integer less than $p - 2$ can be written as a difference of two primes $q - q'$, where q and q' are both less than or equal to p. Due to the concentration of primes at the beginning of the positive numbers, the first 23 odd primes 3, 5, 7, 11, ..., 89 are all cluster primes. The smallest non-cluster prime is 97: the previous prime is 89 and so $88 = 97 - 9$ is not a difference of two primes smaller than 98. In general, if p is a cluster prime, then there must be enough primes in a "small" neighborhood to the left of p so that the even numbers $p - 9$, $p - 15$, $p - 21$, $p - 25$, etc., can all be written as the difference of primes less than p.

The notion of cluster primes is reminiscent of the notion of prime constellations. The most famous prime constellation is that of the twin primes $\{p - 2, p\}$. More elaborate constellations such as $\{p - 8, p - 6, p - 2, p\}$ have also been studied [**4**, pp. 64–68]. In a prime constellation both the number of primes and the differences between them are fixed. There is no guarantee that the largest prime in a prime constellation is a cluster prime. A sparse set of primes may lie just in front of this constellation. The first pair of twin primes $\{p - 2, p\}$ for which p is *not* a cluster prime is $\{227, 229\}$, because the number $p - 27 = 202$ is not a difference of primes less than 230. To see this, observe that $202 = 211 - 9 = 223 - 21 = 227 - 25$, and 211, 223, 227 (the twin of 229) are the only primes between 202 and 228. We show in the proof of Theorem 1 that the number of primes in a small interval just before a cluster prime p grows in size with p. In this way, a cluster prime can be thought of as the largest prime in a "galaxy" of primes.

It is reasonable to expect that among the primes, the cluster primes become increasingly rare. In spite of the initial head start of 23 consecutive cluster primes, the non-cluster primes quickly catch up, so that by the prime 2251 we have 167 cluster primes and 167 non-cluster primes. Starting with 2267, the next prime after 2251, the cluster primes begin to lag further and further behind. When we reach 10^{13}, the non-cluster primes outnumber the cluster primes by a ratio of about 325 to 1.

The simplest question we can ask about the distribution of cluster primes is:

Are there infinitely many cluster primes?

An affirmative answer would imply that $p_{n+1} - p_n \leq 6$ for infinitely many primes p_n, which is a well-known hopeless problem. We enjoy more success looking for an *upper* bound for $\pi_c(x)$, the number of cluster primes not exceeding x. In Section 2 we show that eventually $\pi_c(x)$ is less than $x/(\log x)^s$, for any fixed positive integer s. Our result "suggests" that the cluster primes are less numerous than the twin primes, although we have no way of proving that either of these two collections is infinite. Our theorem is powerful enough, however, to show that the sum of the reciprocals of the cluster primes converges, a result well-known for the twin primes.

In Section 3 we present an efficient algorithm, which, given a particular cluster prime p_n, determines the next cluster prime greater than p_n. We used this algorithm to compute the cluster primes up to 10^{13} and we give the values of $\pi_c(x)$ for powers $x = 10^k$, where k ranges from 2 to 13. These data and a discussion of the results are presented in Section 4. We conclude with a comparison of the number of cluster primes versus the number of pairs of twin primes.

2. AN UPPER BOUND FOR $\pi_c(x)$. We have the following result:

Theorem 1. *For every positive integer s, there is a bound $x_0 = x_0(s)$ such that if $x \geq x_0$ then*

$$\pi_c(x) < \frac{x}{(\log x)^s}. \tag{1}$$

The proof is based on the following two lemmas:

Lemma 1. *Let $\pi(x)$ denote the number of primes $\leq x$. Then for $x \geq 6$,*

$$\pi(x) < \frac{2x - 6}{\log x}.$$

Proof: We use the estimate of Rosser and Schoenfeld [**5**]:

$$\pi(x) < \frac{1.256x}{\log x}, \qquad x > 1. \tag{2}$$

Since

$$\frac{2x - 6}{\log x} > \frac{1.256x + (0.7x - 6)}{\log x} > \frac{1.256x}{\log x}$$

for all $x \geq 9$, Lemma 1 follows for $x \geq 9$. One can easily verify that the formula in the lemma holds for $6 \leq x < 9$. ■

Our main tool in proving Theorem 1 is the following application of Brun's sieve. The notation $f(x) \ll g(x)$ means that for some constant M and value x_0, $|f(x)| \leq Mg(x)$ for all $x \geq x_0$.

Lemma 2. *Let s be a natural number, let $d_1, \ldots, d_s$ be s distinct, nonzero integers, and let $f(x)$ count the number of primes p in the interval $0 < p \leq x$ such that the differences $p - d_i$ are prime for each $i = 1, \ldots, s$. Then*

$$f(x) \ll \prod_{p \mid \prod_{1 \leq i < j \leq s}(d_i - d_j)} \left(1 - \frac{1}{p}\right)^{\rho(p) - s} \prod_{p \mid d_1 \cdots d_s} \left(1 - \frac{1}{p}\right)^{-1} \frac{x}{(\log x)^{s+1}},$$

where $\rho(p)$ denotes the number of modulo p distinct numbers among the d_i's, and where the constant implied by the $\ll$-notation depends only on s.

For a proof of this lemma, take $y = x$ in Corollary 2.4.2 in [**2**, p. 81].

Proof of Theorem 1: Suppose p is a cluster prime. We wish to get a lower bound on the number of primes in the interval $[p - t, p)$, where t is a small positive integer, to be specified later. By the definition of cluster prime, every even number $2r$ in the interval $p - t \leq 2r \leq p - 3$ must be of the form $q - q'$, where q and q' are primes $\leq p$. Clearly q' must be $\leq t$. By Lemma 1, if $t \geq 6$ the number of these primes q' is less than $2(t - 3)/\log t$. On the other hand, there are more than

$(t-3)/2$ even numbers in the interval $[p-t, p-3]$. Thus there must be at least $\frac{1}{4}\log t$ primes in $[p-t, p)$. Define

$$s = \left\lfloor \frac{\log t}{4} \right\rfloor. \tag{3}$$

There are $\binom{t}{s}$ ways to place s numbers $q_1 > q_2 > \cdots > q_s$ in the interval $[p-t, p)$. (We allow the q_i to be even to simplify the calculations.) Using the crude estimate $\binom{t}{s} \le nt^s$, there are fewer than t^s choices for the s differences $d_i = p - q_i$, $1 \le i \le s$. If the differences $d_1 < \cdots < d_s$ are fixed, Lemma 2 ensures that the number of choices of $p \le x$ such that each $p - d_i$ is prime is at most $Mx/(\log x)^{s+1}$, where M is a constant depending only on s. Thus

$$\pi_c(x) < M\frac{x}{(\log x)^{s+1}}t^s.$$

Now given s, let t be the least positive integer satisfying equation (3). Taking x so large that $t^s \le \log x$, we have

$$\pi_c(x) < M\frac{x}{(\log x)^{s}}.$$

Theorem 1 follows easily. ∎

A consequence of Theorem 1 is the following result.

Theorem 2. *The sum of the reciprocals of the cluster primes is finite.*

Proof: If the set of cluster primes is finite, there is nothing to prove, so assume there are infinitely many, and denote the n-th cluster prime by q_n. Consider Theorem 1 with $s = 2$. For n sufficiently large,

$$\pi_c(q_n) < \frac{q_n}{(\log q_n)^2}.$$

But $\pi_c(q_n) = n$ and $(\log q_n)^2 > (\log n)^2$. Thus, for n sufficiently large, we have $q_n > n(\log n)^2$. Since the series $\sum n^{-1}(\log n)^{-2}$ converges, by the integral test, it follows that $\sum 1/q_n$ converges by the comparison test. ∎

It appears that a stronger result than Theorem 1 may actually be true:

Conjecture. *For some constant α, we have*

$$\pi_c(x) \ll \frac{x}{e^{\alpha(\log\log x)^2}}. \tag{4}$$

This result would follow from Lemma 2 if we could guarantee that the implied constant in the lemma does not grow too fast as a function of s.

3. THE ALGORITHM. Let p_n be the nth prime. We describe an algorithm that inputs the index n of the current cluster prime p_n and returns the index of the next cluster prime. The idea is simple. Since p_n is a cluster prime, we know that every even integer from 2 to $p_n - 3$ can be expressed as a difference of two primes not greater than p_n. In order to check whether the next prime p_{n+1} is also a cluster prime, we need examine only the even numbers $p_n - 1$ through $p_{n+1} - 3$. If $p_{n+1} - p_n = 2$, 4, or 6, then there is nothing to check; p_{n+1} is the next cluster prime. If $p_{n+1} - p_n \ge 8$, then p_{n+1} is a non-cluster prime, since $p_{n+1} - 9$ cannot be a difference of two smaller primes. In this case we examine the even numbers

$p_{n+1} - t$, where t is an odd composite less than or equal to $p_{n+1} - p_n + 1$. For each such t we look ahead in the sequence of primes $\{p_{n+m+1}\}_{m=1}^{\infty}$ until $q' = t + p_{n+m+1} - p_{n+1}$ turns out to be prime. Here lies the significance of little m (as opposed to capital M) in the algorithm: p_{n+m+1} is the *first* prime for which the *particular* even number $p_{n+1} - t$ can be written as a difference $p_{n+1+m} - q'$ of primes; so we are at least m primes away from the next cluster prime at this stage of the algorithm. Capital M is the maximum number of primes to the next (possible) cluster prime, based on all of the previous values of m found so far. When we move on to the *next* prime at the end of the outer do-loop, we decrease M by 1, since we are now one prime closer to the cluster prime we are seeking. We continue processing consecutive primes until $M = 0$, indicating that we have finally reached the next cluster prime. For example, the next prime after the non-cluster prime $p_{25} = 97$ is $p_{26} = 101$. Since $p_{25} - 9 = p_{26} - 13$, 101 is the next cluster prime after 89. In our algorithm, we need a list of the prime differences $\text{diff}[n] = p_{n+1} - p_n$. We do not require the actual values of the primes themselves, just the differences, since

$$p_{n+m+1} - p_{n+1} = \sum_{i=1}^{m} \text{diff}[n+i].$$

We also need a short list, named odd_comp, of the odd composites 9, 15, 21, 25, 27, 33, etc., as well as a look-up table to tell when a "small" odd integer is prime.

Algorithm Find_next_cluster_prime(n, current_prime).

```
M = 0;
do
  d = diff[n];
  if (d > 6)
    for (i = 1; odd_comp[i] ≤ d + 1; i = i + 1)
      m = 0;
      t = odd_comp[i];
      repeat
        m = m + 1;
        t = t + diff[n + m];
      until t is prime;
      if (m > M) M = m;
  n = n + 1;
  current_prime = current_prime + d;
  M = M - 1;
while (M > 0)
return n;
```

It is worth pointing out that the efficiency of this algorithm is due to the fact that it always looks forward and *never needs to backtrack*. In actual practice, the program spends more time sieving for the prime differences than it does running the algorithm.

4. DISTRIBUTION OF CLUSTER PRIMES UP TO 10^{13}. We encoded this algorithm in a C program and ran it on a 300 MHz Sun Ultra 2 Workstation. Our goal was to tabulate the cluster primes up to 10^{13} in order to get an indication of their distribution. The following short table gives the number of cluster primes versus non-cluster primes for powers of 10. Here $\pi_c(x)$, $\pi_n(x)$, and $\pi_2(x)$, respectively,

denote the number of cluster primes, non-cluster primes, and twin prime constellations less than or equal to x. Since we do not count 2 as a cluster or non-cluster prime, we have the equation $\pi_c(x) + \pi_n(x) + 1 = \pi(x)$, which we can use as a check on the data. The values of $\pi_2(x)$ were computed by Brent [1] and can also be found in [2, p. 262]. The last column gives the value of α for which (4) becomes an equality, viz. $\alpha = \log(x/\pi_c(x))/(\log\log x)^2$.

x	$\pi_c(x)$	$\pi_n(x)$	$\frac{\pi_n(x)}{\pi_c(x)}$	$\pi_2(x)$	α
10^2	23	1	0.04	8	0.6301
10^3	99	68	0.69	35	0.6192
10^4	420	808	1.92	205	0.6430
10^5	1807	7784	4.31	1224	0.6722
10^6	8287	70,210	8.47	8169	0.6952
10^7	40,017	624,561	15.61	58,980	0.7144
10^8	202,208	5,559,246	27.49	440,312	0.7308
10^9	1,059,812	49,787,721	46.98	3,424,506	0.7455
10^{10}	5,736,857	449,315,654	78.32	27,412,679	0.7586
10^{11}	31,914,282	4,086,140,530	128.03	224,376,048	0.7707
10^{12}	182,065,897	37,425,846,120	205.56	1,870,585,220	0.7817
10^{13}	1,061,375,739	345,004,161,099	325.05	15,834,664,872	0.7921

By $x = 10^4$ the non-cluster primes outnumber the cluster primes by a ratio of roughly 2 to 1. As expected, the ratio $\pi_n(10^k)/\pi_c(10^k)$ increases for each exponent k, and when we reach 10^{13}, approximately 0.3% of the primes are cluster primes. We can show this behavior by using Theorem 1 together with the prime number theorem, which states that $\pi(x)$, the number of primes not greater than x, is asymptotic to $x/\log x$. Since $\pi_c(x)$ is eventually less than $x/(\log x)^2$, it follows that the ratio $\pi_c(x)/\pi(x)$ approaches 0 as x tends to infinity. Thus, "most" primes are non-cluster primes and the ratio $\pi_n(x)/\pi_c(x) = (\pi(x) - 1)(\pi_c(x))^{-1} - 1$ must tend to infinity.

It is interesting to contrast the columns for π_n and π_c. The ratios $\pi_n(10^{k+1})/\pi_n(10^k)$ seem to be approaching the limit 10. A proof of this observation follows from the fact that $\pi_n(x)$ is asymptotic to $\pi(x)$ and from the prime number theorem:

$$\lim_{k\to\infty} \frac{\pi_n(10^{k+1})}{\pi_n(10^k)} = \lim_{k\to\infty} \frac{\pi(10^{k+1})}{\pi(10^k)} = \lim_{k\to\infty} \frac{10^{k+1}}{(k+1)\log 10} \Big/ \frac{10^k}{k\log 10} = 10.$$

For the cluster primes, the ratios $\pi_c(10^{k+1})/\pi_c(10^k)$ increase from 4.24 for $k = 3$ to 5.83 for $k = 12$. It is difficult to predict a limit from such limited data. Heuristic considerations, however, suggest that $\pi_c(x)$ has the shape $x^{1-h(x)}$, where $h(x)$ is a function whose limit tends to 0 as x goes to infinity. If this estimate is correct, then the ratios $\pi_c(10^{k+1})/\pi_c(10^k)$ would also tend to 10, though more slowly than the ratios for the non-cluster primes.

As the program ran to 10^{13}, the largest value of M in the algorithm Find_next_cluster_prime was 58 and the largest value of t was 1503. The largest number of consecutive non-cluster primes was 10,543, found between the cluster primes 8,353,771,390,333 and 8,353,771,707,107. The difference between these two primes is 316,774, the largest difference found up to 10^{13}.

It is worthwhile noting that past 10^6, the number of cluster primes lags behind the number of twin primes. For $x = 10^{12}$ the number of twin primes is roughly ten times larger than the number of cluster primes. To explain this phenomenon, put

$s = 1$ and $d_1 = 2$ into Lemma 2 (Brun's sieve) to get the upper bound

$$\pi_2(x) \ll \frac{x}{(\log x)^2}. \tag{5}$$

Brun used this estimate in 1921 to show that the sum of the reciprocals of the twin primes converges; the proof is essentially the same as our proof of Theorem 2. Comparing (5) with Theorem 1's estimate $\pi_c(x) \ll x/(\log x)^s$ for any positive integer s, we would expect the cluster primes to be rarer than the twin primes. In the interest of honesty, however, we must admit two facts. First, the estimate in (2) for $\pi_c(x)$ holds for $x \geq x_o(s)$. On examining the proof of Theorem 1, the value of $x_0(s)$ is roughly $x_0(s) = e^{t^s}$, where $t = e^{4s}$. For $s = 3$, this bound is $x_0 = e^{e^{36}}$, a number with approximately 1.87×10^{15} decimal digits. It seems a bit presumptuous to think that we are seeing the effects of Theorem 1 with $s = 3$ for the comparatively small 13 digit numbers. The second remark is that although upper bounds may indicate what happens, they are not conclusive. For all we know, the number of cluster primes and twin primes could both be finite. In 1922 Hardy and Littlewood conjectured that $\pi_2(x)$ is asymptotic to

$$2\prod_{p\geq 3}\frac{p(p-2)}{(p-1)^2}\int_2^x \frac{dx}{(\log x)^2} \approx 1.320323632\int_2^x \frac{dx}{(\log x)^2}.$$

This famous conjecture has been shown to be remarkably accurate in estimating the number of twin prime constellations [**3**, p. 66], and strengthens our belief that $\pi_2(x)arrow\infty$ with x. That $\pi_c(x)$ tends to infinity seems harder to prove.

ACKNOWLEDGMENTS. We thank Heini Halberstam for helping elucidate the phrase "by Brun's sieve" in Paul Erdős' handwritten proof of Theorem 1 in Section 2. We are also grateful for discussions with David Rusin. Michael Filaseta is responsible for many improvements throughout the manuscript. Finally, Eric Behr helped with the systems requirements for implementing the cluster prime algorithm.

REFERENCES

1. R. P. Brent, Irregularities in the distribution of primes and twin primes, *Math. Comp.* **29** (1975) 43–56.
2. H. Halberstam and H. E. Richert, *Sieve Methods*, Academic Press, 1974.
3. Paulo Ribenboim, *The New Book of Prime Number Records*, Springer, 1996.
4. Hans Riesel, *Prime Numbers and Computer Methods for Factorization*, Birkhäuser, 1985.
5. J. Barkley Rosser and L. Schoenfeld, Approximate formulas for some functions of prime numbers, *Illinois J. Math.* **6** (1962) 64–97.

RICHARD BLECKSMITH grew up in Cincinnati, Ohio and has been a longtime fan of baseball (Reds), basketball (Bulls), motorcycles (Harleys), and mathematics (Number Theory). He earned his PhD from the University of Arizona in 1983 under the direction of John Brillhart and is a professor at Northern Illinois University. With his wife Sharon, daughter Katie and various itinerate mathematicians, he lives on a small farm near Malta, Illinois, where he is often out standing in his field. He is Secretary of the Number Theory Foundation and is also Vice President of the Malta Band Boosters.
Department of Mathematical Sciences, Northern Illinois University, DeKalb, Illinois 60115
richard@math.niu.edu

J. L. SELFRIDGE was born and raised in Ketchikan, Alaska. He earned his BS in Math at the University of Washington and his PhD at U.C.L.A. He found factors of the Fermat numbers F_{16} and F_{10} in 1953 and has solved various Erdős problems for 40 years, culminating in about 15 joint papers. Pomerance, Selfridge, and Wagstaff offer $620 for a counter-example or proof of an easy primality test. (See page 28 of Richard Guy's *Unsolved Problems in Number Theory*, 2nd edition.) His life included 15 exciting years of full-time administration. Retirement since 1991 has been a mathematical feast.
Department of Mathematical Sciences, Northern Illinois University, DeKalb, Illinois 60115
selfridg@math.niu.edu

NOTES

Edited by **Jimmie D. Lawson and William Adkins**

Fermat's Last Theorem for Gaussian Integer Exponents

John A. Zuehlke

In this note, we observe that Wiles' Theorem [2] on the impossibility of

$$x^n + y^n = z^n$$

for x, y, z positive rational numbers with integer exponents $n \neq \pm 1, \pm 2$ can be generalized to the case of Gaussian integer exponents $\nu = n + im$ without additional exceptions. The proof uses the Gelfond-Schneider Theorem [1], according to which α^β is transcendental for β algebraic but not rational and α algebraic $\neq 0, 1$.

The proof almost fits into the margin. In fact, from

$$x^\nu + y^\nu = z^\nu, \text{ with } \nu = n + im, \quad m \neq 0$$

it follows by taking the complex modulus squared that

$$x^{2n} + 2x^n y^n \cos\theta + y^{2n} = z^{2n}, \text{ with } \theta = m \log(x/y),$$

so $\cos\theta$ is rational. Since, for any real number θ whatsoever there is the identity

$$e^{2i\theta} - 2\cos\theta e^{i\theta} + 1 = 0,$$

it follows for the particular θ that

$$e^{i\theta} = (x/y)^{im}$$

is algebraic. Then the Gelfond-Schneider Theorem, with $\alpha = x/y$ and $\beta = im$ forces $x = y$. Therefore

$$(z/x)^\nu = 2,$$

forcing $z = x$ similarly, contradicting $y \neq 0$.

We remark that the generalization holds, with the same proof, for exponents $\nu = n + im$, with n an integer and m a real algebraic number.

REFERENCES

1. A. Baker, *Transcendental Number Theory*, Cambridge University Press (2nd edition), Cambridge 1979.
2. A. Wiles, Modular Elliptic Curves and Fermat's Last Theorem, *Annals of Math.* **141** (1995) 443–551.

Columbia University, New York, New York 10027
jaz@cpw.math.columbia.edu

On Rational Function Approximations to Square Roots

M. J. Jamieson

Interest in methods for calculating square roots exists partly because the speed with which a computer can evaluate them contributes a measure of its overall performance [**1**]. Newton's well known method for improving estimates of a square root uses a simple rational function approximation and converges in second order; an iterative method converges in second order if it generates a sequence $\{s_n\}$ that tends to limit s and, in which the error $(s_{n+1} - s)$ tends to $K(s_n - s)^2$ for some K, independent of n, as n tends to infinity. This note presents a rational function approximation method with faster convergence. It uses a result of Frank [**2**] on periodic continued fractions; a formula given in the 15th century by Al-Kalsadi [**3**, p. 111] is also based on continued fractions and is a special case.

Frank studied properties of the convergents (or approximants) to the pure periodic continued fraction representing the quadratic surd

$$L + \sqrt{C} = \left[b_0, \overline{b_1, \ldots, b_p}\right] = b_0 + \cfrac{1}{b_1 + \cfrac{1}{b_2 + \cfrac{1}{b_3 + \cdots}}}, \tag{1}$$

where L and C are rational numbers with C positive, the b's are positive integers, p is the period and the overbar indicates the periodic part of the continued fraction. There are restrictions on the values of L and C in order that the representing continued fraction be pure periodic [**4**, p. 101], but they are satisfied if L is zero and C is an integer (when $b_p = 2b_0$, although this is not important here). A special case of a property given by Frank ([**2**, (2.2)] with $s = 1$, $C = N$, $L = 0$) is

$$x_{i+j} = (x_i x_j + N)/(x_i + x_j) \quad \text{for } i, j > 0, \tag{2}$$

where x_i is the $(pi - 1)$th convergent to the continued fraction. It can be shown by induction, from Pascal's triangle rule for binomial coefficients $\binom{i}{j}$, that the convergents satisfy

$$x_k = F_k(x_1) \quad \text{for } k > 0, \tag{3}$$

where, if k is odd $(= 2m + 1)$,

$$F_k(x) = \sum_{i=0}^{i=m} \binom{k}{k-2i} x^{k-2i} N^i \div \sum_{i=0}^{i=m} \binom{k}{k-2i-1} x^{k-2i-1} N^i$$

and, if k is even $(= 2m)$,

$$F_k(x) = \sum_{i=0}^{i=m} \binom{k}{k-2i} x^{k-2i} N^i \div \sum_{i=0}^{i=m-1} \binom{k}{k-2i-1} x^{k-2i-1} N^i. \tag{4}$$

The function $F_k(x)$ has a fixed point at $\sqrt{N}$. Equation (3) gives x_k in terms of x_1. By considering the period of the continued fraction to be the multiple $k^n p$ of p instead of p itself we obtain the same formula giving $x_{k^{n+1}}$ in terms of x_{k^n}. Thus function (4) generates a series of approximations to $\sqrt{N}$ which form a subsequence of the convergents to the continued fraction if one starts with $x = x_1$.

We can use function (4) in an iterative scheme for finding the square root of an integer. To find the square root of a rational number q/r, say, we calculate $\sqrt{qr}$ and divide by r.

The theory of continued fractions guarantees convergence (unless $k = 1$) but we must know the value of the $(p - 1)$th convergent to start the iterative sequence. This is inconvenient. However, by the following theorem, convergence is also guaranteed if we start with an arbitrary positive value x.

Theorem. *For* $x > 0$, $y_n := F_{K^n}(x) \to \sqrt{N}$.

We generalize (2) and (3) and replace (2) by

$$F_{i+j}(x) = \left[F_i(x)F_j(x) + N\right] \div \left[F_i(x) + F_j(x)\right] \quad \text{for } i, j > 0. \tag{5}$$

From (3), (4), and (5) we find

$$F_k(x) - \sqrt{N} = \left(x - \sqrt{N}\right)\left[F_{k-1}(x) - \sqrt{N}\right] \div \left[x + F_{k-1}(x)\right] \quad \text{for } k > 1. \tag{6}$$

From the definition of $F_k(x)$, if x is positive so is $F_k(x)$ for any k. By induction from (3), (4), and (6) we find

$$F_k(x) \geq \sqrt{N} \quad \text{if} \quad x \geq \sqrt{N}, \tag{7a}$$

$$F_k(x) < x \quad \text{if} \quad x > \sqrt{N}. \tag{7b}$$

These inequalities imply that repeated application of the function $F_k(x)$ generates a strictly monotonic decreasing sequence whose greatest lower bound is $\sqrt{N}$ if the starting value exceeds $\sqrt{N}$. Hence convergence is guaranteed for any starting value exceeding $\sqrt{N}$. An argument similar to that leading to inequality (7a) shows that

$$F_k(x) \geq \sqrt{N} \quad \text{if} \quad x \leq \sqrt{N} \text{ for } k \text{ even}, \tag{8a}$$

$$F_k(x) \leq \sqrt{N} \quad \text{if} \quad x \leq \sqrt{N} \text{ for } k \text{ odd}. \tag{8b}$$

For even k and starting value smaller that $\sqrt{N}$ the first generated value exceeds $\sqrt{N}$ and convergence is guaranteed by the argument of the preceding paragraph. If k is odd it can be shown from (6) and (8) that the function $F_k(x)$ is strictly monotonic increasing with least upper bound equal to $\sqrt{N}$ for $x \leq \sqrt{N}$. Thus convergence is guaranteed for any positive starting value. Equation (6) shows that the convergence is of order k.

The first two estimates of $\sqrt{2}$ (1.414213562) obtained with

$$F_4(x) = (x^4 + 6x^2N + N^2)/(4x^3 + 4xN) \tag{9}$$

and starting value 1 are $17/12 = 1.416666667$ and $665857/470832 = 1.414213562$, rounded to nine decimal places; convergence is rapid.

The Newton method is

$$F_2(x) = (x^2 + N)/2x. \tag{10}$$

The approximation of Al-Kalsadi is

$$\sqrt{(a^2 + b)} \approx (4a^3 + 3ab)/(4a^2 + b), \tag{11}$$

which is $F_3(a)$ with N replaced by $a^2 + b$; with starting value 1 ($a = 1$) the first estimate of $\sqrt{2}$ is $7/5 = 1.4$.

REFERENCES

1. J. Bentley, Programming pearls—birth of a cruncher, *Communications of the Association for Computing Machinery* **29** (1986) 1155–1161.
2. E. Frank, On continued fractions for binomial quadratic surds, *Numerische Mathematik* **4** (1962) 85–95.
3. F. Cajori, *A history of mathematics*, second edition, Macmillan, New York, 1958.
4. H. Davenport, *The higher arithmetic—an introduction to the theory of numbers*, Hutchinson's University Library, London, 1952.

Department of Computing Science, University of Glasgow
mjj@dcs.glasgow.ac.uk

A Note on Jacobi Symbols and Continued Fractions

A. J. van der Poorten and P. G. Walsh

1. INTRODUCTION. It is well known that the continued fraction expansion of a real quadratic irrational is periodic. Here we relate the expansion for $\sqrt{rs}$, under the assumption that $rX^2 - sY^2 = \pm 1$ has a solution in integers X and Y, to that of $\sqrt{r/s}$ and to the Jacobi symbols $\left(\frac{r}{s}\right)$, which appear in the theory of quadratic residues.

We have endeavoured to make our remarks self-contained to the extent of providing a brief reminder of the background theory together with a cursory sketch of the proofs of the critical assertions. For extensive detail the reader can refer to [**5**], the bible of the subject. The introductory remarks following in Sections 2–3 below are *inter alia* detailed in [**1**].

Let p and q denote distinct odd primes. In [**3**], Friesen proved connections between the value of the Legendre symbol $\left(\frac{p}{q}\right)$ and the length of the period of the continued fraction expansion of $\sqrt{pq}$. These results, together with those of Schinzel in [**6**], provided a solution to a conjecture of Chowla and Chowla in [**2**].

We report a generalization of those results to the evaluation of Jacobi symbols $\left(\frac{r}{s}\right)$, and, in the context of there being a solution in integers X, Y to the equation $rX^2 - sY^2 = \pm 1$, remark on the continued fraction expansion of $\sqrt{r/s}$ *vis à vis* that of $\sqrt{rs}$.

Theorem 1. *Let r and s be squarefree positive integers with $r > s > 1$, such that the equation $rX^2 - sY^2 = \pm 1$ has a solution in positive integers X, Y. Suppose the continued fraction expansion of $\sqrt{rs}$ is $[a_0, \overline{a_1, a_2, \ldots, a_l}]$. Then both the length of the period $l = 2h$, and the 'central' partial quotient a_h, are even, and the continued fraction expansion of $\sqrt{r/s}$ is*

$$\left[\tfrac{1}{2}a_h, \overline{a_{h+1}, \ldots, a_l, a_1, \ldots, a_h}\right] = \left[\tfrac{1}{2}a_h, \overline{a_{h-1}, \ldots, a_1, a_l, a_1, \ldots, a_{h-1}, a_h}\right].$$

Theorem 2. *Let r and s be squarefree positive integers with $r > s > 1$, such that the equation $rX^2 - sY^2 = \pm 1$ has a solution in positive integers X, Y. Denote by l the length of the period of the continued fraction expansion of $\sqrt{rs}$. Then the following Jacobi symbol equalities hold:*

$$\left(\frac{r}{s}\right) = \left(\frac{-1}{s}\right)^{\frac{1}{2}l+1}, \qquad \left(\frac{s}{r}\right) = \left(\frac{-1}{r}\right)^{\frac{1}{2}l}.$$

As an immediate consequence we obtain the following results, which respectively appeared as Theorem 2 and Theorem 5 in [3].

Corollary 1. *Let $p \equiv q \equiv 3 \pmod 4$ be distinct primes and set $N = pq$. Denote by l the length of the period of the continued fraction expansion of $\sqrt{N}$. Then l is even, and*

$$\left(\frac{p}{q}\right) = \epsilon(-1)^{\frac{1}{2}l},$$

where $\epsilon = 1$ if $p < q$ and $\epsilon = -1$ if $p > q$.

Corollary 2. *Let $p \equiv 3 \pmod 8$ and $q \equiv 7 \pmod 8$ be primes and set $N = 2pq$. Denote by l the length of the period of the continued fraction expansion of $\sqrt{N}$. Then l is even, and*

$$\left(\frac{p}{q}\right) = \epsilon(-1)^{\frac{1}{2}l},$$

where $\epsilon = 1$ if $2p < q$ and $\epsilon = -1$ if $2p > q$.

2. CONTINUED FRACTIONS. *In this section we recall some basic facts about continued fractions that will be appealed to in the proof of our results.*

Given an irrational number α, define its sequence $(\alpha_i)_{i \geq 0}$ of complete quotients by setting $\alpha_0 = \alpha$, and $\alpha_{i+1} = 1/(\alpha_i - a_i)$. Here, the sequence $(a_i)_{i>0}$ of *partial quotients* of α is given by $a_i = \lfloor \alpha_i \rfloor$ where $\lfloor \cdot \rfloor$ denotes the integer part of its argument. Plainly we have

$$\alpha = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ddots}}}$$

It is only the partial quotients that matter, so such a continued fraction expansion may be conveniently denoted just by $[a_0, a_1, a_2, a_3, \ldots]$.

The truncations $[a_0, a_1, \ldots, a_i]$ plainly are rational numbers p_i/q_i. Here, the pairs of relatively prime integers p_i, q_i are given by the matrix identities

$$\begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_i & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} p_1 & p_{i-1} \\ q_i & q_{i-1} \end{pmatrix}$$

and the remark that the empty matrix product is the identity matrix. The alleged correspondence, whereby matrix products provide the *convergents* p_i/q_i, may be confirmed by induction on the number of matrices on noticing the definition

$$[a_0, a_1, \ldots, a_i] = a_0 + 1/[a_1, \ldots, a_i], \qquad [a_0] = a_0.$$

Incidentally, it clearly follows from transposing the matrix correspondence that

$$[a_i, a_{i-1}, \ldots, a_1] = q_i/q_{i-1}, \quad \text{for } i = 1, 2, \ldots . \tag{1}$$

The matrix correspondence entails $p_i/q_i = p_{i-1}/q_{i-1} + (-1)^{i-1}/q_{i-1}q_i$ whence, by induction, $\alpha = a_0 + \sum_{i=1}^{\infty}(-1)^{i-1}/q_{i-1}q_i$, and so

$$0 < (-1)^{i-1}(q_i\alpha - p_i) < 1/q_{i+1},$$

displaying the excellent quality of approximation to α provided by its convergents. Conversely, if

$$|q\alpha - p| < 1/2q, \tag{2}$$

then the rational p/q must be a convergent to α.

3. CONTINUED FRACTIONS OF SQUARE ROOTS OF RATIONALS. If $\alpha = \sqrt{N}$, for positive integer N not a square, it is well known and easy to confirm by induction that its complete quotients α_i are all of the shape $\alpha_i = (P_i + \sqrt{N})/Q_i$, with the sequences of integers (P_i) and (Q_i) given sequentially by $P_{i+1} + P_i = a_iQ_i$, and $Q_{i+1}Q_i = N - P_{i+1}^2$, where $\alpha_0 = \sqrt{N}$ entails $P_0 = 0$ and $Q_0 = 1$. Plainly, always $P_i^2 \equiv N \pmod{Q_i}$. Moreover, it is easy to see that the integers P_i all satisfy $0 \le P_i < \sqrt{N}$ and the positive integers Q_i are all less than $2\sqrt{N}$. It follows by the box principle that the continued fraction expansion of $\sqrt{N}$ must be periodic. Much more is fairly clear.

First, note that the generic step in the continued fraction algorithm for $\alpha = \sqrt{N}$ is $\alpha_i = (P_i + \sqrt{N})/Q_i = a_i - (P_{i+1} - \sqrt{N})/Q_i$. Under conjugation $\sqrt{N} \mapsto -\sqrt{N}$, this step transforms to

$$\left(P_{i+1} + \sqrt{N}\right)/Q_i = a_i - \left(P_i - \sqrt{N}\right)/Q_i. \tag{3}$$

But the 0-th step, ingeniously adjusted by adding $a_0 = P_1$,

$$a_0 + \sqrt{N} = 2a_0 - \left(a_0 - \sqrt{N}\right)$$

is plainly invariant under conjugation. Moreover, because $-1 < P_1 - \sqrt{N} < 0$ we have $(P_1 + \sqrt{N})/Q_1 > 1$. On the other hand $P_1 + \sqrt{N} > 1$ of course entails $-1 < (P_1 - \sqrt{N})/Q_1 < 0$. It's now easy to see, by induction on i, that in (3) $-1 < (P_i - \sqrt{N})/Q_i < 0$. So a_i is the integer part of $(P_{i+1} + \sqrt{N})/Q_i$ and (3) is a step in the continued fraction expansion of $a_0 + \sqrt{N}$, and thus of $\sqrt{N}$.

It follows that the sequence of steps detailing the continued fraction expansion of $a_0 + \sqrt{N}$ is inverted by conjugation, that since it has a fixed point the entire tableaux must be periodic, and that, with l the length of the period, we must have

$$a_0 + \sqrt{N} = \left[\overline{2a_0, a_1, a_2, \ldots, a_{l-1}}\right], \tag{4}$$

moreover with the *word* $a_1a_2 \cdots a_{l-1}$ a palindrome.

Lemma. *The symmetry just mentioned entails that for even period length l there is a 'central' step, at $h = \frac{1}{2}l$,*

$$\alpha_h = \left(P_h + \sqrt{N}\right)/Q_h = a_h - \left(P_{h+1} - \sqrt{N}\right)/Q_h,$$

invariant under conjugation. So $P_{h+1} = P_h$, and $a_h = 2P_h/Q_h$. It follows that $Q_h|2N$. Conversely, if N is squarefree and $Q_j|2N$, where $j \neq 0$, then l is even and $2j = l$.

Proof: Plainly $Q_h|N - P_h^2$ and $Q_h|2P_h$ entails $Q_h|2N$. As regards the converse, it suffices to notice that $Q_j|N - P_j^2$ and $Q_j|2N$ implies $Q_j|2P_j^2$. The only possible square factor of Q_j is 4, since N is squarefree and $Q_j|2N$, so it follows that $Q_j|2P_j$; say $2P_j/Q_j = a_j$. Thus, $\alpha_j = (P_j + \sqrt{N})/Q_j = a_j - (P_j - \sqrt{N})/Q_j$ is a step in the continued fraction expansion of $\sqrt{N}$ invariant under conjugation. It therefore must be the central such step, and this is what we were to show. ∎

Again by induction, or otherwise, one can confirm that

$$\begin{pmatrix} p_1 & p_{i-1} \\ q_i & q_{i-1} \end{pmatrix}\begin{pmatrix} 1 & P_{i+1} \\ 0 & Q_{i+1} \end{pmatrix} = \begin{pmatrix} p_1 & Nq_i \\ q_i & p_i \end{pmatrix};$$

which entails in particular that $p_i^2 - Nq_i^2 = (-1)^{i+1}Q_{i+1}$. In other words, the Q_{i+1} arise from the convergents as just indicated.

Conversely, one sees that when $x^2 - Ny^2 = t$ with $|t| < \sqrt{N}$ then, if $t > 0$, $|x/y - \sqrt{N}| < 1/2y^2$, whilst if $t < 0$ then $|y/x - 1/\sqrt{N}| < 1/2x^2$. In either case it follows from the remark following (2) that x/y is a convergent to $\sqrt{N}$, whence t is $(-1)^{i+1}Q_{i+1}$, some i.

4. PROOF OF THE THEOREMS

Proof of Theorem 1. Set $N = rs$. By the definitions of the sequences (P_i) and (Q_i) we have

$$\left(P_h + \sqrt{N}\right)/Q_h = [a_h, a_{h+1}, a_{h+2}, \dots] = \left[a_h, \overline{a_{h+1}, \dots, a_l, a_1, a_2, \dots, a_h}\right].$$

Let (X, Y) be a positive integer solution to $rX^2 - sY^2 = \pm 1$. Then $(sY)^2 - NX^2 = \mp s$, so because $s < \sqrt{N}$ it follows that sY/X is a convergent to $\sqrt{rs}$, and, more to the point, s is some Q_i for $\sqrt{N}$. Since, trivially, $s|N = rs$, we see that the lemma entails that $i = \frac{1}{2}l = h$, whence $sY/X = p_{h-1}/q_{h-1}$.

Here $Q_h = s$ is squarefree by hypothesis and, since now it divides N, the argument given at the lemma entails that $Q_h|P_h$. Thus $P_h/Q_h = \frac{1}{2}a_h$ is an integer, and so

$$\sqrt{rs}\,/Q_h = \sqrt{r/s} = \left[\tfrac{1}{2}a_h, \overline{a_{h+1}, \dots, a_l, a_1, \dots, a_h}\right].$$

Finally, our remark at (1), or, if one prefers, the observation at (4) that the word $a_1a_2 \cdots a_{l-1}$ is a palindrome, provides the given formulation of the expansion. ∎

Proof of Theorem 2. We saw in the proof of Theorem 1 that the data entails $sY/X = p_{h-1}/q_{h-1}$. Thus

$$p_{h-1}^2 - Nq_{h-1}^2 = (-1)^h Q_h \quad \text{is} \quad (sY)^2 - rsX^2 = (-1)^h s,$$

and so $sY^2 - rX^2 = (-1)^h$, from which the desired conclusions follow. ∎

We now establish the proofs of the corollaries.

Proof of Corollary 1. With $N = pq$ divisible by a prime congruent to 3 modulo 4, it is plain that $U^2 - NV^2 = -1$ has no solution in nonzero integers U, V. Thus the period of $\sqrt{pq}$ has even length $l = 2h$, say. Hence there is a solution in relatively prime integers x, y for $x^2 - pqy^2 = \pm Q_h$ with some Q_h dividing $2pq$, and $1 < Q_h < 2\sqrt{pq}$.

However, it is plain that $x^2 - pqy^2 \equiv 2 \pmod 4$ is impossible so Q_h must be one of p or q; say $Q_h = q$. But $x^2 - pqy^2 = \mp q$ implies $x = qY$, $y = X$, giving a solution in integers X, Y to $pX^2 - qY^2 = \pm 1$, satisfying the conditions of Theorem 2. ∎

Proof of Corollary 2. As in the proof of Corollary 1, $U^2 - 2pqV^2 = -1$ is impossible in nonzero integers U, V, so there is a solution in relatively prime integers x, y for $x^2 - 2pqy^2 = \pm Q_h$, for some Q_h dividing $4pq$, and $1 < Q_h < 2\sqrt{2pq}$.

It's easy to see that the possibilities modulo 8 are $x^2 - 2pqy^2 = \pm 2p$ or $x^2 - 2pqy^2 = \pm q$ and that either yields integers X, Y satisfying $2pX^2 - qY^2 = \pm 1$. Thus again the hypotheses of Theorem 2 are satisfied, and the result follows by noticing that the Jacobi symbol $\left(\frac{2}{q}\right) = 1$ for $q \equiv 7 \pmod 8$. ■

5. CLOSING REMARKS. Suppose we know both that

$$\sqrt{r/s} = \left[\tfrac{1}{2}a_h, \overline{a_{h+1}, \ldots, a_l, a_1, \ldots, a_h}\right] \quad \text{and} \quad \sqrt{rs} = \left[a_0, \overline{a_1, \ldots, a_h, a_{h+1}, \ldots, a_l}\right].$$

The two expansions have the same 'tail', that is, they differ only in a finite number of initial partial quotients. Thus the numbers $\sqrt{rs}$ and $\sqrt{r/s}$ are *equivalent* and one sees, for example from the matrix correspondence, that there are integers X, Y, U, and V satisfying $VX - UY = \pm 1$ and so that $(U\sqrt{r/s} + B)(X\sqrt{r/s} + Y) = \sqrt{rs}$. But, removing the surd from the denominator yields

$$\frac{(rUX - sVY) + (VX - UY)\sqrt{rs}}{rX^2 - sY^2} = \sqrt{rs}.$$

It follows that $rUX - sVY = 0$ and, this is the point, $rX^2 - sY^2 = \pm 1$. So the shape of the two continued fraction expansions, and first principles, shows that there is a solution in integers X, Y to $rX^2 - sY^2 = \pm 1$.

We might also recall a cute result mentioned by Nagell [4]. Namely, given an integer N, consider the collection of all equations $aX^2 - bY^2 = \pm 1$ with integers a and b so that $ab = N$. Nagell's remark is that at most two of that collection of diophantine equations can have a solution. One of us happened to have been reminded of this fine fact by Dmitri Mit'kin at a meeting at Minsk, Belarus in 1996.

Proof: The cases $N \leq 0$ or $N = \square$ are uninteresting and trivial, so we suppose that N is positive and is not a square. Then we have at least one equation with a solution, namely $1 \cdot X^2 - NY^2 = 1$. Further, if the length l of the period of $\sqrt{N}$ is odd then also $NX^2 - 1 \cdot Y^2 = 1$ has a solution. If there is some other equation with a solution, say $aX^2 - bY^2 = \pm 1$ with $a > b > 1$, then, as we saw above, $(bY)^2 - NX^2 = \mp b$ so $l = 2h$, $b = Q_h$, and $\mp 1 = (-1)^{h+1}$. Thus there is at most one 'other' equation, and if it has a solution then l is not odd. ■

REFERENCES

1. Enrico Bombieri and A. J. van der Poorten, Continued fractions of algebraic numbers, in *Computational Algebra and Number Theory, Sydney* 1992, Wieb Bosma and Alf van der Poorten eds., Kluwer, 1995, pp. 138–154.
2. P. Chowla and S. Chowla, Problems on periodic simple continued fractions, *Proc. Nat. Acad. Sci. USA* **69** (1972) 37–45.
3. C. Friesen, Legendre Symbols and continued fractions, *Acta Arith.* **59** (1991) 365–379.
4. T. Nagell, On a special class of Diophantine equations of the second degree, *Ark. Mat.* **3** (1954) 51–65.
5. O. Perron, *Die Lehre von den Kettenbruchen*, Chelsea Publishing Company, New York, 1950.
6. A. Schinzel, On two conjectures of P. Chowla and S. Chowla concerning continued fractions, *Ann. Mat. Pura Appl.* **98** (1974) 111–117.

Centre for Number Theory Research, Macquarie University, Sydney 2109, Australia
alf@mpce.mq.edu.au

University of Ottawa, 585 King Edward St., Ottawa, Ontario, Canada K1N-6N5
gwalsh@mathstat.uottawa.ca

THE EVOLUTION OF...

Edited by **Abe Shenitzer**
Mathematics, York University, North York, Ontario M3J 1P3, Canada

The Birth of Literal Algebra

I. G. Bashmakova and G. S. Smirnova
Translated from the Russian by Abe Shenitzer

1. MATHEMATICS IN THE FIRST CENTURIES AD. DIOPHANTUS. The Babylonians developed a kind of numerical algebra. Then came Greek geometric algebra.

The third—very important—stage of the development of algebra began in the first centuries AD and came to an end at the turn of the 17th century. Its beginning was marked by the introduction of *literal symbolism* by Diophantus of Alexandria and its end, by the creation of *literal calculus* in the works of Viète and Descartes. It was then that algebra acquired its own distinctive language, which we use today.

The first century BC was a period of Roman conquests and of Roman civil wars. Both took place in the territories of the Hellenistic states and the Roman provinces and were accompanied by physical and economic devastation. One after another, these states lost their independence. The last to fall was Egypt (30 BC). The horrors of war and the loss of faith in a secure tomorrow promoted the spread of religious and mystical teachings and undermined interest in the exact sciences, and in abstract problems in mathematics and astronomy. In Cicero's dialogue *On the state* one of the participants proposes a discussion of why two Suns were seen in the sky. But the topic is rejected, for "even if we acquired profound insight into this matter, we would not become better or happier."

In the second half of the first century BC mathematical investigations came to a virtual halt and there was an interruption in the transmission of the scientific tradition.

At the beginning of the new era, economic conditions in the Hellenistic countries, now turned Roman provinces, gradually improved, and there was a revival of literature, art, and science. In fact, the 2nd century came to be known as the Greek Renaissance. It was the age of writers such as Plutarch and Lucian and of scholars such as Claudius Ptolemy.

Alexandria continued its role as the cultural and scientific center of antiquity and, in this respect, Rome was never its rival. Nor did it ever develop an interest in the depths of Hellenistic science. As noted by Cicero in his *Tusculanae disputationes*, the Romans, unlike the Greeks, did not appreciate geometry; just as in the case of arithmetic, they stopped at narrow, practical knowledge of this subject.

Translator's note. This article is the third chapter of an essay by I. G. Bashmakova and G. S. Smirnova devoted to the rise and evolution of algebra. The whole essay is being translated by Abe Shenitzer and is being reviewed for publication by the Mathematical Association of America.

The first two sentences of this article were added by the translator.

They had little regard for all of mathematics. Even accounting, surveying, and astronomical observations were left to the Greeks, the Syrians, and other conquered nations. According to Vergil, the destiny of Romans was wise government of the world.

The revival of the Alexandrian school was accompanied by a fundamental change of orientation of its mathematical research. During the Hellenistic period geometry was the foundation of Greek mathematics; algebra had not, as yet, become an independent science but developed within the framework of geometry, and even the arithmetic of whole numbers was constructed geometrically. Now number became the foundation. This resulted in the arithmetization of all mathematics, the elimination of geometric justifications, and the emergence and independent evolution of algebra.

We encounter the return to numerical algebra already in the works of the outstanding mathematician, mechanician, and engineer Heron of Alexandria (1st century AD). In his books *Metrica*, *Geometrica*, and others, books that resemble in many respects our handbooks for engineers, one finds rules for the computation of areas and volumes, solutions of numerical quadratic equations, and interesting problems that reduce to indeterminate equations. In particular, they contain the famous "Heron formula" for the computation of the area of a triangle given its sides a, b, c:

$$S = \sqrt{p(p-a)(p-b)(p-c)},$$

where $p = (a + b + c)/2$. Here the expression under the square root sign is a product of four segments, and thus an expression totally inadmissible in geometric algebra. It is clear that Heron thought of segments as numbers, whose products are likewise numbers.

In his famous book, known under its Arabized name *Almagest*, Claudius Ptolemy, when computing tables of chords, identified ratios of magnitudes with numbers, and the operation of "composition " of ratios—defined in Euclid's *Elements*—with ordinary multiplication.

The new tendencies found their clearest expression in the works of Diophantus of Alexandria, who founded two disciplines: algebra and Diophantine analysis.

We know next to nothing about Diophantus himself. On the basis of certain indirect remarks, Paul Tannery, the eminent French historian of mathematics, concluded that Diophantus lived in the middle of the 3rd century AD. On the other hand, Renaissance scholars who discovered Diophantus' works, supposed that he lived at the time of Antoninus Pius, i.e., approximately in the middle of the 2nd century. The epigram in *Anthologia Palatina* provides the following information: "Here you see the tomb containing the remains of Diophantus, it is remarkable: artfully it tells the measures of his life. God granted him to be a boy for the sixth part of his life, and adding a twelfth part to this, He clothed his cheeks with down; He lit him the light of wedlock after a seventh part, and five years after his marriage He granted him a son. Alas! late-born wretched child; after attaining the measure of half his father's life, chill Fate took him. After consoling his grief by this science of numbers for four years he ended his life. By this device of numbers tell us the extent of his life." A simple computation shows that Diophantus died at the age of 84 years. This is all we know about him.

2. DIOPHANTUS' *Arithmetica*. ITS DOMAIN OF NUMBERS AND SYMBOLISM. Only two (incomplete) works of Diophantus have come down to us. One is his *Arithmetica* (six books out of thirteen; four more books in Arabic, attributed to

Diophantus, were found in 1973. They will be discussed in the sequel). The other is a collection of excerpts from his treatise *On polygonal numbers*. We discuss only the first of these works.

Arithmetica is not a theoretical work resembling Euclid's *Elements* or Apollonius' *Conic sections* but a collection of (189) problems, each of which is provided with one or more solutions and with relevant explanations. At the beginning of the first book there is a short algebraic introduction, which is basically the first account of the foundations of algebra. Here the author constructs the field of rational numbers, introduces literal symbolism, and gives rules for operating with polynomials and equations.

Already Heron regarded positive rational numbers as legitimate numbers (in classical ancient mathematics "number" denoted a collection of units, i.e., a natural number). While Diophantus defined a number as a collection of units, throughout *Arithmetica* he called every positive rational solution of one of his problems "number" (ἀριθμός), i.e., he extended the notion of number to all of $\mathbf{Q}^+$. But this was not good enough for the purposes of algebra, and so Diophantus took the next decisive step of introducing negative numbers. It was only then that he obtained a system closed under the four operations of algebra, i.e., a field.

How did Diophantus introduce these new objects? Today we would say that he used the axiomatic method: he introduced a new object called "deficiency" (λεῖψις, from λείπω—to lack) and stated rules for operating with it. He writes: "deficiency multiplied by deficiency yields availability (i.e., a positive number (*the authors*)); deficiency multiplied by availability yields deficiency; and the symbol for deficiency is $\curlywedge$, an inverted and shortened (letter) ψ" (Diophantus. *Arithmetica*. Definition IX). In other words, he formulated the rule of signs, which we can write as follows:

$$(-) \times (-) = (+),$$
$$(-) \times (+) = (-).$$

Diophantus did not formulate rules for addition and subtraction of the new numbers but he used them extensively in his books. Thus, while solving problem III_8 (i.e., Problem 8 in Book III), he needs to subtract $2x + 7$ from $x^2 + 4x + 1$. The result is $x^2 + 2x - 6$, i.e., here he carries out the operation $1 - 7 = -6$. In problem VI_{14}, $90 - 15x^2$ is subtracted from 54 and the result is $15x^2 - 36$. Thus here $15x^2$ is subtracted from zero; in other words, Diophantus is using the rule $-(-a) = a$.

We note that Diophantus used negative numbers only in intermediate computations and sought solutions only in the domain of positive rational numbers. A similar situation developed later in connection with the introduction of complex numbers. Initially they were regarded as just convenient symbols for obtaining results involving "genuine," i.e., real, numbers.

Diophantus also introduced literal signs for an unknown and its powers. He called an unknown a "number" (ἀριθμός) and denoted it by the special symbol ς. It is possible that this symbol was introduced before him. We find it in the Michigan papyrus (2nd century AD) as well as in a table appended to Heron's *Geometrica*. But Diophantus boldly breaks with geometric algebra by introducing special symbols *for the first six positive powers of the unknown, the first six negative powers, and for its zeroth power*. While the square and cube of the unknown could be interpreted geometrically, its 4th, 5th, and 6th powers could not be so represented. Nor could the negative powers of the unknown.

Diophantus denoted the positive powers of the unknown as follows:

$$x\text{—}\varsigma; \quad x^2\text{—}\Delta^\nu; \quad x^3\text{—}K^\nu; \quad x^4\text{—}\Delta^\nu\Delta; \quad x^5\text{—}\Delta K^\nu; \quad x^6\text{—}K^\nu K.$$

He defined negative powers as inverses of the corresponding positive powers and denoted them by adding to the exponents of the positive powers the symbol χ. For example, he denoted $x^{-2} = 1/x^2$ by $\Delta^{\upsilon\chi}$.

He denoted the zeroth power of the unknown by the symbol $\mathring{M}$, that is by the first two letters in Μόνάς, or unity.

Then he set down a "multiplication table" for powers of the unknown that can be briefly written as follows:

$$x^m x^n = x^{m+n}, \quad -6 \le m + n \le 6.$$

He singled out two rules that correspond to the two basic axioms that we use for defining a group:

$$x^m \cdot 1 = x^m \quad \text{(definition VII)}; \tag{1}$$

$$x^m x^{-m} = 1 \quad \text{(definition VI)}. \tag{2}$$

In addition, Diophantus used the symbol $\overline{\iota\sigma}$ for equality, and the symbol □ for an indeterminate square. All this enabled him to write equations in literal form. Since he did not use a symbol for addition, he first set down all positive terms, then the minus sign (i.e., $\pitchfork$), then the negative terms. For example, the equation

$$x^3 - 2x^2 + 10x - 1 = 5$$

was written as

$$\mathrm{K}^{\upsilon}\bar{\alpha}\varsigma\bar{\iota} \pitchfork \Delta^{\upsilon}\bar{\beta}\,\mathring{\mathrm{M}}\,\bar{\alpha}\overline{\iota\sigma}\,\mathring{\mathrm{M}}\,\bar{\epsilon}.$$

Here $\bar{\alpha} = 1$, $\bar{\iota} = 10$, $\bar{\beta} = 2$, $\bar{\epsilon} = 5$ (we recall that the Greeks used the letters of the alphabet to denote numbers).

In the "Introduction" Diophantus formulated two basic rules of transformation of equations: 1) the rule for transfer of a term from one side of an equation to the other with changed sign and 2) reduction of like terms. Later, these two rules became well known under their Arabized names of *al − jabr* and *al − muqābala*.

Diophantus also used the rule of substitution in a masterly way but never formulated it.

We can say that in the introduction Diophantus defined the field **Q** of rational numbers, introduced symbols for an unknown and its powers, as well as symbols for equality and for negative numbers.

Before discussing the contents of *Arithmetica* we consider the possibilities and limitations of Diophantus' symbolism. Getting ahead of the story, we can say that, basically, Diophantus considered in his work indeterminate equations, i.e., equations with two or more unknowns. But he introduced symbols for just one unknown and its powers. How did he proceed when solving problems?

First he stated each problem in general form. For example: "To decompose a square into a sum of squares" (problem II_8). Now we would write this problem as

$$x^2 + y^2 = a^2.$$

How could Diophantus write this equation with just one symbol for an unknown and without symbols for parameters (in this case a)? He proceeded as follows: after the general formulation he assigned concrete values to the parameters—in the present case he put $a^2 = 16$. Then he denoted one unknown by his special symbol (we will use the letter t instead) and expressed the remaining unknowns as linear, quadratic, or more complex rational functions of that unknown and of the parameters. In case of the present example, one unknown is denoted by t and the other by $kt - a$ or, as Diophantus puts it, "a certain number of t's minus as many

units as are contained in the side of 16," i.e., instead of a he takes 4 and instead of the parameter k—the number 2. But by saying "a certain number of t's" he indicates that the number 2 plays the role of an arbitrary parameter. Thus Diophantus' version of our equation is

$$t^2 + (2t - 4)^2 = 16,$$

so that

$$x = t = 16/5; \quad y = 2t - 4 = 12/5.$$

One might think that Diophantus was satisfied with finding a single solution. But this is not so. In the process of solving problem III_{19} he finds it necessary to decompose a square into two squares. In this connection he writes: "We know that a square can be decomposed into a sum of squares in infinitely many ways."

The use of a concrete number to denote an arbitrary parameter has the virtue of simplicity. Sometimes it turned out that the parameter could not be selected arbitrarily, that it had to satisfy additional conditions. In such cases Diophantus determined these conditions. Thus problem VI_8 reduces to the system

$$x_1^3 + x_2 = y^3, \quad x_1 + x_2 = y.$$

Diophantus puts $x_2 = t$, $x_1 = \beta t$, where $\beta = 2$. Then from the second equation we obtain $y = (\beta + 1)t$, and from the first

$$t^2 = \frac{1}{(\beta + 1)^3 - \beta^3}.$$

Since $\beta = 2$, $t^2 = 1/19$, i.e., t is not rational. In order to obtain a rational solution Diophantus looks at the way t^2 is expressed in terms of the parameter β. The expression in question is a fraction whose numerator, 1, is a square. But then the denominator must also be a square:

$$(\beta + 1)^3 - \beta^3 = \square.$$

Diophantus took as the new unknown $\beta = \tau$ (he denoted it by the same symbol as the original unknown x_2) and obtained

$$3\tau^2 + 3\tau + 1 = \square.$$

Solving this equation by his method (which we will describe in detail in the next section) Diophantus obtained

$$\tau = \frac{3 + 2\lambda}{\lambda^2 - 3},$$

i.e., the parameter could only be chosen from the class of numbers $\{(3 + 2\lambda)/(\lambda^2 - 3)\}$. Diophantus takes $\lambda = 2$ and obtains $\beta = 7$. Then he goes back to solving the original problem.

Diophantus often deliberately chooses for parameters numbers that do not lead to solutions. He does this in order to show how to analyze problems.

Thus concrete numbers play two roles in *Arithmetica*. One role is that of ordinary numbers and the other is that of symbols for arbitrary parameters. Numbers were destined to play the latter role almost to the end of the 16th century.

Time to sum up. Diophantus was first to reduce determinate and indeterminate problems to equations. We may say that for a large class of problems of arithmetic and algebra he did the same thing that Descartes was later to do for problems of geometry, namely he reduced them to setting up and solving algebraic equations.

Indeed, in order to solve problems—arithmetical in the case of Diophantus and geometric in the case of Descartes—both set up algebraic equations that they subsequently transformed and solved in accordance with the rules of algebra. Also, the transformations involved— such as elimination of unknowns, substitutions, and reduction of similar terms—had no direct arithmetic or geometric significance and were not subject to extensive interpretations. In both cases such interpretations were reserved for the final results. We are used to associating this important step with Descartes' creation of analytic geometry, but long before Descartes this step was taken by Diophantus in his *Arithmetica*. Before exposure to *Arithmetica*, none of the scholars of the period between the 13th and 16th centuries entertained the idea of applying algebra to the solution of number-theoretic problems.

3. THE CONTENTS OF *Arithmetica*. DIOPHANTUS' METHODS. We have already mentioned that *Arithmetica* is a collection of problems with solutions. This may create the impression that it is not a theoretical work. But a more careful reading makes it clear that the purpose of the painstaking choice and deliberate placement of problems was to illustrate the application of specific general methods. It is a characteristic of ancient mathematics that methods were not formulated apart from problems but were disclosed in the process of their solution. We recall that the famous "method of exhaustion"—the first variant of the theory of limits—was not set down in pure form either by its author Eudoxus of Cnidus or by Archimedes. It was mathematicians of the 16th and 17th centuries who isolated it by analyzing Euclid's *Elements* and Archimedes' quadratures and formulated it in general terms. The same applies to Diophantus' *Arithmetica*. As we show in the sequel, his methods were isolated in the 16th and 17th centuries by Italian and French mathematicians. Following them, we will try to isolate some of these methods and state them in general form.

In Book I Diophantus solved particular determinate linear and quadratic equations. The remaining books deal with the solution of indeterminate equations, i.e., equations of the form

$$F(x_1, \ldots, x_n) = 0, \quad n \geq 2,$$

$F(x_1, \ldots, x_n)$ a polynomial, or of systems of such equations:

$$\begin{cases} F_1(x_1, \ldots, x_n) = 0; \\ \cdots\cdots\cdots\cdots\cdots \\ F_k(x_1, \ldots, x_n) = 0, \quad k < n. \end{cases}$$

Diophantus looks for positive rational solutions $x_1^0, x_2^0, \ldots, x_n^0$, $x_i^0 \in \mathbf{Q}^+$, of such equations or of such systems.

It is clear that to solve his determinate equations Diophantus needed only symbols for x and x^2 and not for x^n, $-6 \leq n \leq 6$. In other words, he extended his domain of numbers and introduced most of his symbols to investigate and solve indeterminate equations, where he really needed higher powers of the unknown as well as its negative powers.

Thus the birth of literal algebra was connected not with determinate but with indeterminate equations.

Here we present just one of Diophantus' methods, namely his method for finding the rational solutions of a quadratic equation in two unknowns:

$$F_2(x, y) = 0, \tag{3}$$

where $F_2(x, y)$ is a quadratic polynomial with rational coefficients.

Basically, Diophantus proves the following theorem: if equation (3) has a rational solution (x_0, y_0) then it has infinitely many such solutions (x, y), and x and y are both rational functions (with rational coefficients) of a single parameter:

$$x = \varphi(k), \quad y = \psi(k). \tag{4}$$

When presenting his methods we use modern algebraic symbolism. This is by now a standard procedure in historical-mathematical literature.

Diophantus began by considering quadratic equations of the form

$$y^2 = ax^2 + bx + c, \quad a, b, c \in \mathbf{Q}, \tag{5}$$

and put $c = m^2$ (in other words, he assumed that the equation had two rational solutions $(0, m)$ and $(0, -m)$). To find solutions he made the substitution

$$y = kx \pm m \tag{6}$$

and obtained

$$x = \frac{b \mp 2km}{k^2 - a}, \qquad y = \frac{b \mp 2km}{k^2 - a} \pm m.$$

By assigning to k all possible rational values (Diophantus took only values that yielded positive x and y) we obtain infinitely many solutions of equation (5).

We note that the substitutions (6) are the famous Euler substitutions that are applied to integrals of the form

$$\int \frac{dx}{\sqrt{ax^2 + bx + c}}.$$

We mentioned earlier problem II_8, which reduces to the equation

$$x^2 + y^2 = a^2, \tag{7}$$

and recall that Diophantus solved it by making the substitution

$$x = t; \quad y = kt - a, \tag{8}$$

and obtained (we are replacing his numerical values by appropriate letters)

$$x = t = a\frac{2k}{1 + k^2}; \qquad y = a\frac{k^2 - 1}{k^2 + 1}.$$

To see the sense of this solution and to appreciate its generality we must look at its geometric interpretation. Equation (7) determines a circle of radius a centered at the origin, and the substitution (8) is the equation of a straight line with slope k passing through the point $A(0, -a)$ on that circle (Figure 1). It is clear that the straight line (8) intersects the circle (7) in another point B with rational coordinates. Conversely, if there is a point B_1 with rational coordinates (x_1, y_1) on the circle (7) then AB_1 is a straight line of the pencil (8) with rational slope k. Thus to every rational k there corresponds a rational point on the circle (7) and to every rational point on the circle (7) there corresponds a rational value of k. Hence Diophantus' method yields all rational solutions of equation (7).

This argument shows that a conic with a rational point is birationally equivalent to a rational straight line.

Next Diophantus considered the more general case when equation (5) has a rational point but the coefficient c is not a square. He first considered this case in problem II_9, which reduces to the equation

$$x^2 + y^2 = a^2 + b^2 \tag{9}$$

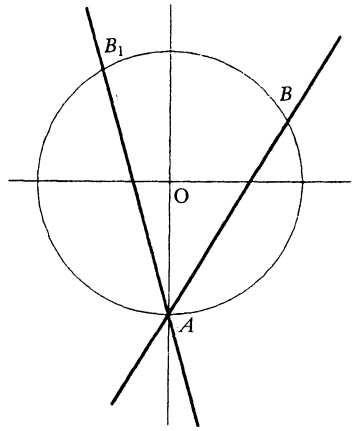


Figure 1

(Diophantus put $a = 2$, $b = 3$). It is clear that equation (9) has the following four solutions: (a, b), $(-a, b)$, $(a, -b)$, and $(-a, -b)$. Diophantus makes the substitution

$$x = t + a, \qquad y = kt - b \tag{10}$$

and obtains $t = 2(bk - a)/(1 + k^2)$. Applying a geometric interpretation analogous to the one just used we see that, essentially, he is leading a straight line with slope k through $(a, -b)$ on the circle (9).

Diophantus considered a more general case in lemma 2, proposition VI_{12} and in the lemma for proposition VI_{15}: assuming that equation (5) has a rational solution (x_0, y_0) he made the substitution $x = t + x_0$ and obtained

$$y^2 = at^2 + (2ax_0 + b)t + y_0^2,$$

i.e., he reduced the problem to the case $c = m^2$.

Finally, he considered equation (5) in the case when $a = \alpha^2$. He made the substitution (easily recognized as Euler's "second substitution" (*the authors*))

$$y = \alpha t \pm k \tag{11}$$

and obtained

$$t = \frac{c - k^2}{2\alpha k - b}.$$

This case calls for a separate discussion. To understand why the straight line (11) intersects the conic section (5) in just one point we introduce projective coordinates (U, V, W) by putting $x = U/W$, $y = V/W$, i.e., we consider our conic in the projective plane P^2. Then equation (5) takes the form

$$\alpha^2 U^2 + bUW + cW^2. \tag{12}$$

The curve L so defined intersects the line at infinity $W = 0$ in two rational points: $(1, \alpha, 0)$ and $(1, -\alpha, 0)$. The straight line (11), whose equation in projective coordinates is

$$V = \alpha U + kW,$$

passes through the first of these points.

In summary, we can say that Diophantus carried out a complete investigation of a quadratic indeterminate equation in two unknowns. Later, his analysis served as a model for the investigation of the question of rational points on curves of genus 0.

Diophantus used more complex and more sophisticated methods to solve equations of the form

$$y^2 = ax^3 + bx^2 + cx + d,$$
$$y^3 = ax^3 + bx^2 + cx + d,$$
$$y^2 = ax^4 + bx^3 + cx^2 + dx + f,$$

and systems of the form

$$\begin{cases} ax^2 + bx + c = y^2, \\ a_1x^2 + b_1x + c_1 = z^2, \end{cases}$$

which he called "double equalities." Readers interested in getting a deeper understanding of Diophantus' methods should consult the book by Bashmakova and Slavutin: *A history of Diophantine analysis from Diophantus to Fermat* (Russian) which contains further references to the literature. The history of Diophantus' methods extends all the way to the papers of Poincaré that appeared at the beginning of the 20th century. It was on the basis of these methods that Poincaré constructed an arithmetic of algebraic curves—an area of mathematics that is being intensively developed at the present time.

We conclude our survey by considering Diophantus' problem III_{19}. This problem reduces to a system of 8 equations in 12 unknowns:

$$\begin{cases} (x_1 + x_2 + x_3 + x_4)^2 + x_i = y_i^2, \\ (x_1 + x_2 + x_3 + x_4)^2 - x_i = z_i^2; \quad i = 1, 2, 3, 4. \end{cases}$$

Diophantus notes that "in every right triangle the square of the hypotenuse remains a square if we add to it, or subtract from it, twice the product of its legs." This means that he must find four right triangles with the same hypotenuse. Indeed, let the sides of the four triangles be a_i, b_i, c, $i = 1, 2, 3, 4$. Then it suffices to put $x_1 + x_2 + x_3 + x_4 = ct$, $x_i = 2a_ib_it^2$, $i = 1, \ldots, 4$. Thus the problem reduces to finding a number c that can be written as a sum of two squares in four different ways. Diophantus solves this essentially number-theoretic problem as follows: he takes two right triangles with sides 3, 4, 5 and 5, 12, 13 respectively, and multiplies the sides of each of them by the hypotenuse of the other. As a result he obtains two right triangles with the same hypotenuse: 39, 42, 65 and 25, 60, 65. Now $5 = 1^2 + 2^2$ and $13 = 2^2 + 3^2$. Using the rule for composition of forms $u^2 + v^2$ known already to the Babylonians, namely

$$\begin{aligned} (u^2 + v^2)(\alpha^2 + \beta^2) &= (\alpha u - \beta v)^2 + (\alpha v + \beta u)^2 \\ &= (\alpha u + \beta v)^2 + (\alpha v - \beta u)^2, \end{aligned}$$

he obtains

$$65 = 5 \cdot 13 = (1^2 + 2^2)(2^2 + 3^2) = 4^2 + 7^2 = 8^2 + 1^2.$$

Using Euclid's formulas for the general solution of $x^2 + y^2 = z^2$ (i.e., $z = p^2 + q^2$; $x = p^2 - q^2$; $y = 2pq$) we obtain two more right triangles with hypotenuse 65: 33, 56, 65 and 63, 16, 65. This completes the solution of the problem.

In connection with this problem Fermat stated that a prime of the form $4n + 1$ could be written as a sum of squares in just one way. Then he gave a formula for the determination of the number of ways in which a given number can be written as a sum of squares. Thus problems involving indeterminate equations led to number-theoretic insights.

Did Diophantus know the theorems formulated by Fermat? It is possible that he did. Jacobi offered a reconstruction of Diophantus' conjectured proofs, but the answer to this question remains hypothetical.

One can hardly overestimate the significance of Diophantus' *Arithmetica* for the subsequent history of algebra. It is no exaggeration to say that its role was comparable to the role of Archimedes' treatises in the history of the differential and integral calculus. We will see that it was the starting point for all mathematicians up to Bombelli and Viète, and that its importance for number theory and for indeterminate equations can be traced up to the present.

4. ALGEBRA AFTER DIOPHANTUS. The period from the 4th to the 6th centuries AD was marked by the precipitous decline of ancient society and learning. But eminent commentators, such as Theon of Alexandria (second half of the 4th century) and his daughter Hypatia (murdered in 418 by a fanatical Christian mob), were still active. In the 5th century there was an exodus of scholars from Alexandria to Athens. Finally, in the 6th century, Eutocius and Simplicius, the last of the great commentators, were expelled from Athens and settled in Persia.

We can turn to the question of the Arabic translations of four books attributed to Diophantus. An analysis of these books, translated at the end of the 9th century from Greek to Arabic by Costa ibn Luca (i.e., the Greek Constantin, son of Luca) shows that they are a reworked version of Diophantus' *Arithmetica*. They contain problems, possibly due to Diophantus, as well as extensive additions and commentaries to them. According to Suidas' Byzantine dictionary, Hypatia wrote commentaries on *Arithmetica*. It is therefore very likely that the four books translated into Arabic are books edited and provided with commentaries by Hypatia. These books contain no new methods, but the material is presented in a complete and systematic manner. Their author went beyond Diophantus by introducing the 8th and 9th powers of the unknown.

The subsequent development of mathematics, including that of algebra, was connected with the Arabic East. Scholars from Syria, Egypt, Persia, and other regions conquered by the Arabs wrote scientific treatises in Arabic.

BIBLIOGRAPHY

1. I. G. Bashmakova, *Diophantus and Diophantine equations*. Transl. by A. Shenitzer. Under review for publication by the Mathematical Association of America.
2. Th. L. Heath, *Diophantus of Alexandria*. New York, Dover, 1964.
3. O. Ore, *Number theory and its history*. New York, McGraw-Hill, 1948.
4. J. H. Silverman and J. Tate, *Rational points on elliptic curves*. New York, Springer, 1992.
5. B. L. van der Waerden, *Science awakening*. Transl. by A. Dresden. New York, Wiley, 1963.
6. A. Weil, *Number theory: An approach through history*. Basel, Birkhäuser, 1983.

PROBLEMS AND SOLUTIONS

Edited by **Gerald A. Edgar, Daniel H. Ullman,** and **Douglas B. West**

with the collaboration of Paul T. Bateman, Mario Benedicty, Paul Bracken, Duane M. Broline, Ezra A. Brown, Richard T. Bumby, Glenn G. Chappell, Randall Dougherty, Roger B. Eggleton, Ira M. Gessel, Bart Goddard, Jerrold R. Griggs, Douglas A. Hensley, Richard Holzsager, John R. Isbell, Robert Israel, Kiran S. Kedlaya, Murray S. Klamkin, Fred Kochman, Frederick W. Luttmann, Frank B. Miles, Richard Pfiefer, Leonard Smiley, John Henry Steelman, Kenneth Stolarsky, Richard Stong, Charles Vanden Eynden, and William E. Watkins.

Proposed problems and solutions should be sent in duplicate to the MONTHLY *problems address on the inside front cover. Submitted problems should include solutions and relevant references. Submitted solutions should arrive at that address before June 30, 1999; Additional information, such as generalizations and references, is welcome. The problem number and the solver's name and address should appear on each solution. An acknowledgement will be sent only if a mailing label is provided. An asterisk (*) after the number of a problem or a part of a problem indicates that no solution is currently available.*

PROBLEMS

10704. *Proposed by Wiliam G. Spohn, Jr., Ellicott City, MD.* Show that there are infinitely many pairs $((a, b, c), (a', b', c'))$ of primitive Pythagorean triples such that $|a - a'|$, $|b - b'|$, and $|c - c'|$ are all equal to 3 or 4. Examples include $((12, 5, 13), (15, 8, 17))$ and $((77, 36, 85), (80, 39, 89))$.

10705. *Proposed by D. W. Brown, Marietta, GA.* A topological space has the *fixed point property* if every continuous function from the space to itself has a fixed point. Is there a countably infinite Hausdorff space with the fixed point property?

10706. *Proposed by James G. Propp, University of Wisconsin, Madison, WI.* Given a finite sequence $(a_1, \ldots, a_n)$, define the *derived sequence* $(b_1, \ldots, b_{n+1})$ by $b_i = s - a_{i-1} - a_i$, where $s = \min_{1 \le i \le n+1}(a_{i-1} + a_i) + \max_{1 \le i \le n+1}(a_{i-1} + a_i)$ and where we interpret both a_0 and a_{n+1} as 0. Let S_0 be the sequence (1) of length 1, and for $n \ge 1$ define S_k to be the derived sequence obtained from S_{k-1}. Thus $S_1 = (1, 1)$, $S_2 = (2, 1, 2)$, $S_3 = (3, 2, 2, 3)$, and $S_4 = (5, 3, 4, 3, 5)$. Show that the middle term of S_{2n} is a perfect square.

10707. *Proposed by John Isbell, State University of New York, Buffalo, NY.* Show that
(**a**) no vector space over an infinite field is a finite union of proper subspaces; and
(**b**) no vector space over an n-element field is a union of n or fewer proper subspaces.

10708. *Proposed by the Western Maryland College Problems Group, Westminster, MD.* Let

$$f(x) = \frac{1}{4}\int_0^\pi \frac{1}{t}\log\left(\frac{1-\cos(x+t)}{1-\cos(x-t)}\right)dt$$

for $x \in (0, \pi)$.
(**a**) Find the Fourier sine series for f.
(**b**) Find the L^2 norm of f.
(**c**) Find $\lim_{x \to 0} f(x)$.

10709. *Proposed by Zoltán Sasvári, Technical University of Dresden, Dresden, Germany.* Let X be a standard normal random variable, and choose $y > 0$. Show that

$$e^{-ay} < \frac{Pr(a \le X \le a+y)}{Pr(a-y \le X \le a)} < e^{-ay+(1/2)ay^3}$$

when $a > 0$. Show that the reversed inequalities hold when $a < 0$.

10710. *Proposed by Bogdan Suceava, Michigan State University, East Lansing, MI.* Let ABC be an acute triangle with incenter I, and let D, E, and F be the points where the circle inscribed in ABC touches BC, CA, and AB, respectively. Let M be the intersection of the line through A parallel to BC and DE, and let N be the intersection of the line through A parallel to BC and DF. Let P and Q be the midpoints of DM and DN, respectively. Prove that A, E, F, I, P, and Q are on the same circle.

SOLUTIONS

When O-H-I Is Isosceles

10547 [1996, 695]. *Proposed by Dan Sachelarie, ICCE Bucharest, and Vlad Sachelarie, University of Bucharest, Bucharest, Romania.* In the triangle ABC, let O be the circumcenter, H the orthocenter, and I the incenter. Prove that the triangle OHI is isosceles if and only if

$$\frac{a^3+b^3+c^3}{3abc} = \frac{R}{2r}.$$

Solution by Walther Janous, Ursulinengymnasium, Innsbruck, Austria. We denote by MPV the reference D. S. Mitrinović, J. E. Pečarić, and V. Volenec, *Recent Advances in Geometric Inequalities,* Kluwer, 1989. Neither IO nor HI is ever as large as HO [MPV, p. 288], so the only way triangle IHO can be isosceles is if $IO = HI$. Also $IO^2 = R^2 - 2Rr$ [MPV, p. 279] and $HI^2 = 4R^2 + 4Rr + 3r^2 - s^2$ [MPV, p. 280], where s is the semiperimeter. Hence $HI = IO$ if and only if $R^2 - 2Rr = 4R^2 + 4Rr + 3r^2 - s^2$. This rearranges to $2s(s^2 - 3r^2 - 6Rr)/12Rrs = R/2r$, or, using $abc = 4Rrs$ [MPV, p. 52] and $a^3+b^3+c^3 = 2s(s^2 - 3r^2 - 4Rr)$ [MPV, p. 52], to $(a^3 + b^3 + c^3)/3abc = R/2r$.

Editorial comment. Another condition equivalent to $HI = IO$, given in problem E2282 [1971, 196; 1972, 397] from this MONTHLY, is that ABC has one angle equal to 60°.

Solved also by J. Anglesio (France), R. Barbara (Lebanon), F. Bellot Rosado (Spain), C. W. Dodge, J. S. Frame, Z. Franco, M. S. Klamkin (Canada), W. W. Meyer, V. Mihai (Canada), C. R. Pranesachar (India), B. Prielipp, V. Schindler (Germany), I. Sofair, M. Tabaâ (Morocco), T. V. Trif (Romania), M. Vowe (Switzerland), GCHQ Problems Group (U. K.), and the proposers.

The Divisible Differences Property

10553 [1996, 809]. *Proposed by Bjorn Poonen, Mathematical Sciences Research Institute, Berkeley, CA, Jim Propp, Massachusetts Institute of Technology, Cambridge, MA, and Richard Stong, Rice University, Houston, TX.* Say that a sequence $\langle q \rangle = q_1, q_1, q_2, \ldots$ of integers has the *divisible differences property* if $(n-m)|(q_n - q_m)$ for all n and m.
(a) Show that if $\langle q \rangle$ has the divisible differences property and $\limsup |q_n|^{1/n} < e - 1$, then there is a polynomial Q such that $q_n = Q(n)$.
(b) Show that there is a sequence $\langle q \rangle$ that has the divisible differences property and satisfies $\limsup |q_n|^{1/n} \le e$, for which q_n is not given by a polynomial in n.
(c)* Is it true that $\limsup |q_n|^{1/n} \ge e$ for all non-polynomial $\langle q \rangle$ with the divisible differences property?

Solution of parts **(a)** *and* **(b)** *by the GCHQ Problem Solving Group, Cheltenham, U. K.* Let $L_n = \text{lcm}\{1, \ldots, n\}$. We need three facts.

Fact 1: $\log(L_n) \sim n$ as $n \to \infty$.

Proof: If p is prime and $p \leq n$, then $p^{\lfloor \log_p(n) \rfloor}$ is the highest power of p that divides n. Therefore, $L_n = \prod_{p \leq n} p^{\lfloor \log_p(n) \rfloor} < \prod_{p \leq n} p^{\log_p(n)} = \prod_{p \leq n} n = n^{\pi(n)}$. Thus, $\log(L_n) < \pi(n) \log n \sim n$, the latter following from the Prime Number Theorem. Conversely, if $1/2 < r < 1$, then $L_n > \prod_{n^r < p \leq n} p > \prod_{n^r < p \leq n} n^r = n^{r(\pi(n) - \pi(n^r))}$. Taking logarithms and using the Prime Number Theorem yields $\log(L_n) > r(\pi(n) - \pi(n^r)) \log n \sim r\left(\frac{n}{\log n} - \frac{n^r}{r \log n}\right) \log n = rn - n^r \sim rn$. Since r is arbitrary, $\log(L_n) \sim n$. □

Fact 2: Given $q_0, q_1, \ldots, q_{n-1}$, the choices for q_n such that $(n-m) | (q_n - q_m)$ for $0 \leq m < n$ lie in the same congruence class modulo L_n.

This is a consequence of the Chinese Remainder Theorem.

Fact 3: If $q_0, q_1, \ldots, q_{n-1}$ has the divisible differences property, and if r_n is the value found by fitting a minimum degree polynomial to $q_0, q_1, \ldots, q_{n-1}$ and extrapolating, then $r_n = \sum_{i=1}^{n} (-1)^{i+1} \binom{n}{i} q_{n-i}$ and $(n-m) | (r_n - q_m)$ for $0 \leq m < n$.

Proof: The formula for r_n is just a restatement of the property that the nth difference of $q_0, q_1, \ldots, q_{n-1}$ is 0. We prove the rest by induction on n, it being trivial when $n = 1$.

The first difference of $q_0, q_1, \ldots, q_{n-1}$ is a sequence of length $n-1$ with the divisible differences property, and its polynomial extrapolation is $r_n - q_{n-1}$. By the inductive hypothesis, $m | (r_n - q_{n-1}) - (q_{n-m} - q_{n-m-1})$ for $m > 0$, and by the hypothesis on $q_0, q_1, \ldots, q_{n-1}$ we have $m | (q_{n-1} - q_{n-m-1})$. Hence $m | (r_n - q_{n-m})$, so we may assume $m = 0$. By subtracting q_0 from the rest of the sequence, we may assume $q_0 = 0$, so

$$r_n = \sum_{i=1}^{n-1} (-1)^{i+1} \binom{n}{i} q_{n-i}. \tag{1}$$

Now $n | (n-i)\binom{n}{i}$ since $(n-i)\binom{n}{i} = n\binom{n-1}{i}$, and $(n-i) | q_{n-i}$, so n divides each term of the sum in (1) and therefore r_n. □

We can restate the divisible differences property as: $L_n | (q_n - r_n)$ for all n.

(a) If $\limsup |q_n|^{1/n} < e - 1$, then there is an $\epsilon > 0$ and an integer n_0 such that $|q_n|^{1/n} < e - 1 - 3\epsilon$ for $n > n_0$. From this and Fact 3, there is an $n_1 > n_0$ such that $|r_n| < \sum_{i=n_0}^{n} \binom{n}{i} (e - 1 - 2\epsilon)^{n-i} < (e - 2\epsilon)^n < (e - \epsilon)^n / 2$ for $n > n_1$. Using Fact 1, there is an $n_2 > n_1$ such that $n > n_2$ implies that $\log(L_n) > n \log(e - \epsilon)$ (that is, $L_n > (e - \epsilon)^n$). If $n > n_2$ and $q_0, q_1, \ldots, q_{n-1}$ has the divisible differences property, r_n is thus the only value for q_n that would extend the property and also have absolute value less than $(e - \epsilon)^n / 2$. Since $|q_n| < (e - 1 - 3\epsilon)^n$, it follows that $q_n = r_n$.

This is the inductive step in a proof that the minimum degree polynomial that fits $q_0, \ldots, q_{n_2}$ fits q_n for all n.

(b) Set $q_0 = 0$. For each positive n, set q_n to be congruent to r_n mod L_n but of opposite sign and having magnitude less than L_n. If $r_n = 0$, however, set $q_n = L_n$. Since q_n never equals r_n, q_n is not given by a polynomial. Since $|q_n| \leq |L_n|$, we have $\limsup |q_n|^{1/n} \leq e$.

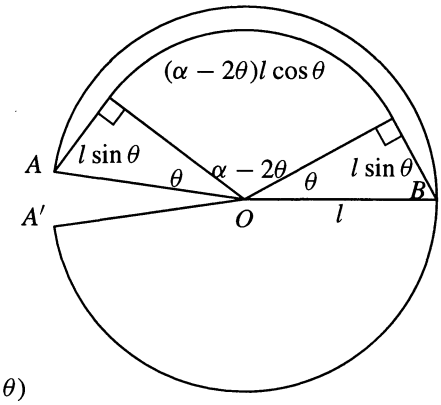
Editorial comment. The results of parts **(a)** and **(b)** appear in I. Ruzsa, On congruence preserving functions (in Hungarian), *Mat. Lapok* 22 (1971) 125-134 (*Math. Reviews*, vol. 48, #2044). Part **(a)** generalizes problem 4 from the 1995 USA Mathematical Olympiad, which imposed the stronger condition that q_n is bounded by a polynomial. No solutions were submitted for part **(c)**.

Parts **(a)** and **(b)** solved also by R. J. Chapman (U. K.), K. S. Kedlaya, and the proposers.

Mathematics for Hikers

10557 [1996, 902]. *Proposed by Nick MacKinnon, Winchester College, Winchester, U. K.* Naismith's rule allows walkers to compute the time for their journeys. The time is given by allowing a walking speed of 4 km/hr, but adding an extra minute for each 10m of ascent. A conical mountain has base radius 1650m and vertical height 520m. Points A and B are diametrically opposite at the base of the mountain. How should a path be constructed between between A and B on the surface of the mountain that minimizes the time taken to walk from A to B?

Solution by the proposer. The surface of the mountain can be unrolled isometrically onto a plane, forming a sector (of angle $2\alpha = 165/173 \cdot 2\pi$ radians) of a circle (of radius $l = 1730$m, the slant height of the mountain), as shown in the figure at right.



Call the time-minimising path the *Naismith geodesic* for the cone. This geodesic must reach some maximum height h. The figure shows a potential Naismith geodesic with maximum height $h = 520(1 - \cos\theta)$ meters. It is composed of a circular arc following the contour with height h, together with tangents joining the arc to A and B. No alternative path taking less time attains the maximum height h of the given path since such a path must at least touch the circular arc, must not cross the circular arc, and must leave the circular arc without subsequent reascent. The Naismith geodesic must therefore be a path of the given shape. The length of such a path is $2l\sin\theta + (\alpha - 2\theta)l\cos\theta$, and Naismith's rule gives a time of

$$t(\theta) = 0.015(2l\sin\theta + (\alpha - 2\theta)l\cos\theta) + 520(1 - \cos\theta)/10$$

minutes for the journey. The lone critical value of $t(\theta)$ for $0 < \theta < \alpha/2$ occurs when $t'(\theta) = (52 - 0.015(\alpha - 2\theta)l)\sin\theta = 0$ at $\theta^* = (\alpha - 52/(0.015l))/2 \doteq .49623$ radians. This gives the optimal time $t(\theta^*) = 76.71037$ minutes. A path around the bottom of the cone takes $t(0) = 77.75442$ minutes, while the path of shortest distance takes $t(\alpha/2) = 99.98929$ minutes.

Solved also by J. Anglesio (France), J. E. Dawson (Australia), P. G. Kirmser, J. H. Lindsey II, R. Reynolds & M. Martinez, and P. Strafflin.

A Matter of Adjustment

10558 [1996, 902]. *Proposed by Zhang Chengyu, Hubei University, Wuhan, China.* Let p be a prime, and let k be a positive integer. Let $a_1, a_2, \ldots, a_{p^k}$ be any p^k integers. We define the *adjustment* of these integers to be the p^k integers $b_1, b_2, \ldots, b_{p^k}$, where $b_j = a_{j+1} + a_{j+2} + \cdots + a_{j+p}$, interpreting subscripts modulo p^k. For example, if $p = 2$ and $k = 2$, one adjustment of 1, 1, 3, 4 gives 4, 7, 5, 2. Prove that after p^k adjustments of $a_1, a_2, \ldots, a_{p^k}$, the list consists entirely of integers divisible by p.

Solution I by Thomas Jager, Calvin College, Grand Rapids, MI. We prove a stronger statement. Given an integer vector $v = (v_1, \ldots, v_{p^k})$, define the v-adjustment of $(a_1, \ldots, a_{p^k})^T$ to be $(b_1, \ldots, b_{p^k})^T$, where $b_j = v_1 a_{j+1} + v_2 a_{j+2} + \cdots + v_{p^k} a_{j+p^k}$, again treating subscripts modulo p^k. As a transformation, the v-adjustment is represented by the matrix $A = v_1 S + v_2 S^2 + \cdots + v_{p^k} S^{p^k}$, where S is the permutation matrix for a cyclic shift by

one position. Hence

$$A^{p^k} = \left(v_1 S + \cdots + v_{p^k} S^{p^k}\right)^{p^k} \equiv v_1^{p^k} I + \cdots + v_{p^k}^{p^k} I \equiv (v_1 + \cdots + v_{p^k})^{p^k} I \pmod{p}.$$

Thus if $v_1 + \cdots + v_{p^k} \equiv 0$ modulo p, then p^k applications of the v-adjustment matrix produces a vector of integers divisible by p. In the problem statement, the vector v consists of p ones and $p^k - p$ zeros.

Solution II by J. H. van Lint, Eindhoven University of Technology, Eindhoven, the Netherlands. Starting with a_0, we construct an infinite sequence with $a_i = a_{i+p^k}$. Over the field $\mathbb{F}_p$, we consider the formal power series $f(x) = \sum_{i=0}^{\infty} a_i x^i = A(x)/(1 - x^{p^k})$, where $A(x) = \sum_{i=0}^{p^k-1} a_i x^i$ is a polynomial of degree less than p^k.

After one adjustment, the terms $b_0, b_1, \ldots$ are the coefficients of $x^{p+1}, x^{p+2}, \ldots$ in the formal power series for

$$(x + x^2 + \cdots + x^p) f(x) = \frac{x(1 - x^p)}{1 - x} f(x) = x(1 - x)^{p-1} f(x).$$

The result of n adjustments is the list of coefficients of $x^{n(p+1)}, x^{n(p+1)+1}, \ldots$ in the formal power series for

$$x^n (1 - x)^{n(p-1)} f(x) = \frac{x^n (1 - x)^{n(p-1)}}{1 - x^{p^k}} A(x),$$

which is a polynomial of degree less than np if $n(p-1) \geq p^k$. Thus the list consists entirely of integers divisible by p after n adjustments if $n \geq p^k/(p-1)$. Noting that

$$\frac{p^k - 1}{p - 1} + 1 = \frac{p^k}{p - 1} + \frac{p - 2}{p - 1}$$

is the least integer greater than or equal to $p^k/(p-1)$, we see that the list consists entirely of integers divisible by p after n adjustments if $n \geq (p^k - 1)/(p - 1) + 1$. As this is at most p^k, the desired result follows.

Editorial comment. David Callan proved that for positive m the list consists entirely of integers divisible by p^{m-1} after mp^{k-1} adjustments. In particular, after p^k adjustments the list consists entirely of integers divisible by p^{p-1}. Another consequence is that the list consists entirely of integers divisible by p after $2p^{k-1}$ adjustments, but this is not as strong as the result proved by van Lint.

Solved also by D. Beckwith, A. E. Caicedo Núñez (Colombia), D. Callan, R. J. Chapman (U. K.), J. E. Dawson (Australia), W. Janous (Austria), K. S. Kedlaya, J. H. Lindsey II, R. Martin (Germany), A. Nijenhuis, J. C. Smith, H.-T. Wee (Singapore), GCHQ Problems Group (U. K.), and the proposer.

Sets with Fixed Nim-Sum

10564 [1997, 68]. *Proposed by Proposed by Aviezri Fraenkel, Weizmann Institute of Science, Rehovot, Israel.* The *Nim-sum* of two positive integers with binary expansions $\sum_{i \geq 0} a_i 2^i$ and $\sum_{i \geq 0} b_i 2^i$ is the number with binary expansion $\sum_{i \geq 0} c_i 2^i$, where a_i, b_i, c_i are in $\{0, 1\}$ and $c_i \equiv a_i + b_i \bmod 2$. Let n be a positive integer, and let j be a nonnegative integer. How many of the 2^n subsets of $\{1, 2, \ldots, n\}$ have the property that their elements have Nim-sum equal to j?

Solution by Reiner Martin, Deutsche Bank, London, U. K. Let $[n] = \{1, 2, \ldots, n\}$, and let Δ denote the symmetric difference operation. Let $k = \lceil \log_2(n + 1) \rceil$. There exists a subset of $[n]$ whose elements have Nim-sum j only if $0 \leq j < 2^k$. We claim that the number of such subsets does not depend upon j and thus that this number is 2^{n-k} for each such j.

To prove this, let $\sum_{i\geq 0} a_i 2^i$ be the binary expansion of j, and let $A_j = \{2^j : a_j = 1\}$. For each $A \subseteq [n]$ with Nim-sum 0, let $f(A) = A \Delta A_j$. Note that $f(A)$ has Nim-sum j. Since $(A \Delta A_j) \Delta A_j = A$, this map is a bijection into the set of subsets of $[n]$ with Nim-sum j.

Solved also by D. Beckwith, M. Benedicty, D. Berstein, J. C. Binz (Switzerland), M. Bowron, D. Callan, R. J. Chapman (U. K.), D. Donini (Italy), G. Gordon, R. Holzsager, K. S. Kedlaya, N. Komada, J. H. Lindsey II, J. Lorch, O. P. Lossers (The Netherlands), D. K. Nester, A. Nijenhuis, K. O'Bryant, M.-K. Siu (Hong Kong), J. H. Steelman, W. Stromquist, I. Vardi (Canada), H.-T. Wee (Singapore), M. Wolterman, Anchorage Math Solutions Group, GCHQ Problems Group (U. K.), NCCU Problems Group, NSA Problems Group, and the proposer.

Generalized Line Bingo

10565 [1997, 68]. *Proposed by D. M. Bloom, Brooklyn College, Brooklyn, NY, and Kenneth Suman, Winona State University, Winona, MN.* A rectangle is composed of mn squares arranged in m rows and n columns. In a certain game, the squares are selected one by one at random (without replacement). What is the expected number of selections until j columns of the rectangle are composed entirely of selected squares? (When $j = 1$, $m = 5$, and $n = 15$, this is the expected length of a type of bingo game known as a line game.)

Composite solution by the GCHQ Problems Group, Cheltenham, U. K. and the editors. For fixed m and n, the required expectation E_j equals $mn \prod_{i=j}^{n-1} mi/(mi+1)$.

For each instance of the game, we can continue selecting squares at random until all squares are selected. Thus it suffices to compute, over all permutations of the mn squares, the expected length of the initial segment that completes j columns. We compute for each square x the probability that it belongs to that initial segment. This is independent of x, so the expectation is mn times this probability.

Let A_i be the event that x belongs to the initial segment in which i columns are completed; note that $Pr(A_n) = 1$. The probability $Pr(A_j)$ is the product over $i \geq j$ of $Pr(A_i | A_{i+1})$.

We partition A_{i+1} into subevents that fix the trailing segment after the position where the $(i+1)$st column is completed. In such a subevent S, the identities of the first $i+1$ finished columns are fixed, but not which of these is last.

For permutations in S, let B be the set of squares consisting of the first i finished columns and the last square that completes the $(i+1)$st completed column. When $x \in B$, it is equally likely to occupy any of the $mi+1$ positions occupied by B, so the fraction of such permutations that belong to A_i is $mi/(mi+1)$.

When $x \notin B$, we can group the permutations by each fixed permutation of B. Now x is equally likely to fall into each of the $mi+1$ segments between members of B (or before the first). Again the fraction of these permutations that belong to A_i is $mi/(mi+1)$.

Editorial comment. The rows are unimportant. Víctor Hernández used linearity of expectation and the inclusion-exclusion principle to obtain a formula in the more general situation where the columns are sets of arbitrary size.

Solved also by R. J. Chapman (U. K.), D. A. Darling, V. Hernández (Spain), R. Holzsager, J. H. Lindsey II, P. W. Lindstrom, N. C. Singer, J. C. Smith, J. H. Steelman, Anchorage Math Solutions Group, and the proposer.

Ordered Trees and Stirling Numbers

10570 [1997, 69]. *Proposed by Emeric Deutsch, Polytechnic University, Brooklyn, NY.* An ordered tree is a rooted tree in which the children of each node form a sequence rather than a set. The height of an ordered tree is the number of edges on a path of maximum length starting at the root. Let $a(n, k)$ denote the number of ordered trees with n edges and height k, and let $S(n, k)$ be the Stirling number of the second kind (the number of partitions of $\{1, 2, \ldots, n\}$ into k nonempty parts). Note that $a(n, 1) = S(n, 1)$, since both numbers are 1. Show that (**a**) $a(n, 2) = S(n, 2)$, (**b**) $a(n, 3) + a(n, 4) = S(n, 3)$, and (**c**)* generalize these observations.

Solution I by Robin J. Chapman, University of Exeter, Exeter, UK. Let $b(n, k)$ be the number of ordered trees with n edges and height at most k. We include the tree with a root and no edges, so $b(0, k) = 1$ for $k \geq 0$. It suffices to show that $b(n, 2) = 1 + S(n, 2)$ and $b(n, 4) = 1 + S(n, 2) + S(n, 3)$ for $n > 0$. To achieve this, we compute the generating function $g_k(x) = \sum_{n \geq 0} b(n, k)x^n$ for $0 \leq k \leq 4$.

We have $g_0(x) = 1$. For $k > 0$, an ordered tree of height at most k consists of the root v, r edges incident to it, and a sequence of r ordered trees of height at most $k - 1$ rooted at the children of v. The generating function for ordered trees of height at most k in which the root has degree r is thus $x^r(g_{k-1}(x))^r$. Summing over r, we obtain $g_k(x) = 1/(1 - xg_{k-1}(x))$. Explicitly, this yields

$$g_1(x) = \frac{1}{1-x}, \qquad g_3(x) = \frac{1-2x}{1-3x+x^2},$$

$$g_2(x) = \frac{1-x}{1-2x}, \quad \text{and} \quad g_4(x) = \frac{1-3x+x^2}{1-4x+3x^2}.$$

Expanding by partial fractions yields

$$g_2(x) = 1 + \frac{x}{1-2x} \qquad \text{and} \qquad g_4(x) = 1 + \frac{x}{2(1-x)} + \frac{x}{2(1-3x)}.$$

Thus $b(n, 2) = 2^{n-1}$ and $b(n, 4) = (3^{n-1} + 1)/2$ for $n \geq 1$.

The number of partitions of $[n]$ into at most two parts is half the number of subsets of $[n]$, so $1 + S(n, 2) = 2^{n-1}$, as desired. Now consider partitions of $[n]$ into at most three parts. Each element other than n enters the part with n or one of the other two. Thus 3^{n-1} counts each partition with at least two parts twice, as we can interchange the second and third part without changing the partition. The partition with one part appears only once, so the total number of classes is $(3^{n-1} + 1)/2$, as desired.

Finding further relations among these numbers seem unlikely. If $f_k(x) = f_{k-1}(x) - xf_{k-2}(x)$ for $k \geq 2$, with $f_0(x) = f_1(x) = 1$, then $g_k(x) = f_k(x)/f_{k+1}(x)$ for all k. One can show that

$$f_k(x) = \sum_{j=0}^{\lfloor k/2 \rfloor} (-1)^j \binom{k-j}{j} x^j = \prod_{j=1}^{\lfloor k/2 \rfloor} \left(1 - 4x\cos^2 \frac{j\pi}{k+1}\right).$$

It follows that

$$b(n, k) = \sum_{j=1}^{\lfloor (k+1)/2 \rfloor} r_{k,j} \left(4\cos^2 \frac{j\pi}{k+2}\right)^n$$

for constants $r_{k,j}$. On the other hand, it is well known that $S(n, k) = \sum_{j=1}^{k} t_{k,j} j^n$ for constants $t_{k,j}$. When $k \notin \{0, 1, 2, 4\}$, the value $4\cos^2 j\pi/(k + 2)$ is irrational for some j, so there is little hope of establishing a simple relationship betweeen $b(n, k)$ and the Stirling numbers in these cases.

Solution II to parts **(a)** *and* **(b)** *by Daniele Donini, Bertinoro, Italy.* We define a bijection from the set of ordered trees with n edges and height at most 4 to the set of partitions of $[n]$ with at most 3 blocks, in which for $k \leq 3$ the trees with height k become partitions with k blocks, and the trees with height 4 become partitions with 3 blocks.

Given a tree T, label the non-root vertices with the integers 1 through n in order via a depth-first left-first search. To form the corresponding partition, let the ith block consists of the label on vertices at distance i for the root, except as follows. In each subtree rooted at a vertex at distance 3 from the root, put the largest value in block 3 and put the other values in block 1. This reduces to partitioning by levels for trees with height 3.

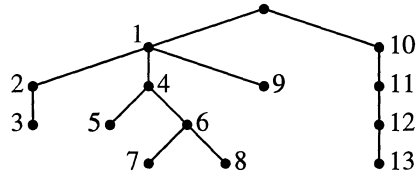
For the inverse map, index the blocks in partition π in increasing order of their least elements: A_1, possibly A_2, possibly A_3. Build the corresponding tree traversal (starting with 1 as the leftmost vertex at level 1) as follows. Suppose that the label just processed was label k, belonging to A_i. Let A_j be the block containing label $k+1$.

Case 1: $j \leq i$. Make $k+1$ a new rightmost leaf at level j.

Case 2: $j = i+1$. Make $k+1$ the first child of the vertex with label k.

Case 3: $j = 3$ and $i = 1$. Because $\min A_2 < \min A_3$, there exists a label less than k in A_2. Let l be the largest such label less than k. Let m be the least label such that all labels from m to k lie in A_1; note that $m > l$. Remove $m, \ldots, k$ from level 1. Make m the rightmost child of l (at level 3). Make $m+1, \ldots, k+1$ children of m (at level 4).

Applying the original map to the resulting tree puts each label back into its block in π. As an example, the tree corresponding to $A_1 = \{1, 6, 7, 10, 12\}$, $A_2 = \{2, 4, 9, 11\}$, and $A_3 = \{3, 5, 8, 13\}$ is shown.



Solved also by D. Callan, R. Holzsager, the Anchorage Math Solutions Group, and the proposer.

Lattice Points Inside a Triangle

10600 [1997, 566]. *Proposed by Franz Rothe, University of North Carolina, Charlotte, NC.*
(a) Suppose a triangle has its vertices at integer lattice points in the plane and contains exactly 3 integer lattice points in its interior. Show that the center of mass of the triangle is not an integer lattice point.
(b)* Find all values i such that, if a triangle has its vertices at integer lattice points in the plane and contains exactly i integer lattice points in its interior, then the center of mass of the triangle cannot be an integer lattice point.

Solution of part **(a)** *by Robin J. Chapman, University of Exeter, Exeter, U. K.* Let the vertices of the triangle be A, B, and C, with position vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$, respectively. Suppose that the centroid of the triangle has integer coordinates. This centroid is $(1/3)(\mathbf{a}+\mathbf{b}+\mathbf{c})$. Let D denote twice the area of the triangle. Then

$$D = |(\mathbf{b}-\mathbf{a})\times(\mathbf{c}-\mathbf{a})| = |(\mathbf{c}-\mathbf{b})\times(\mathbf{a}-\mathbf{b})| = |(\mathbf{a}-\mathbf{c})\times(\mathbf{b}-\mathbf{c})| = |\mathbf{a}\times\mathbf{b}+\mathbf{b}\times\mathbf{c}+\mathbf{c}\times\mathbf{a}|$$

where $\times$ denotes the vector product. By assumption $(1/3)(\mathbf{a}+\mathbf{b}+\mathbf{c})$ has integer coordinates; therefore, so has $(1/3)(\mathbf{a}+\mathbf{b}+\mathbf{c}) - \mathbf{a} = (1/3)\big((\mathbf{b}-\mathbf{a}) + (\mathbf{c}-\mathbf{a})\big)$. Hence $(\mathbf{c}-\mathbf{a}) = -(\mathbf{b}-\mathbf{a}) + 3\mathbf{d}$, where $\mathbf{d}$ has integer coordinates, and so $D = |(\mathbf{b}-\mathbf{a})\times(\mathbf{c}-\mathbf{a})| = 3|(\mathbf{b}-\mathbf{a})\times\mathbf{d}|$ is a multiple of 3. Let r, s, and t be the largest integers such that $(1/r)(\mathbf{b}-\mathbf{c})$, $(1/s)(\mathbf{c}-\mathbf{a})$, and $(1/t)(\mathbf{a}-\mathbf{b})$ have integer coordinates. Then the interiors of the sides BC, CA, and AB contain respectively $r-1$, $s-1$, and $t-1$ lattice points. By Pick's Theorem $D = 2N_1 + N_2 - 2$, where N_1 is the number of lattice points in the interior of the triangle and N_2 is the number of lattice points on its boundary (including the vertices). Consequently $D = 4 + r + s + t$. Also

$$D = st\left|\frac{\mathbf{b}-\mathbf{a}}{t}\times\frac{\mathbf{c}-\mathbf{a}}{s}\right|,$$

and so D is divisible by st. Similarly, D is divisible by rs and rt.

Since D is divisible by 3, we have $r+s+t \equiv 2 \pmod 3$. We now show that none of r, s, t is divisible by 3. Suppose that r is divisible by 3. Then $\mathbf{b}-\mathbf{c} = (\mathbf{b}-\mathbf{a})+(\mathbf{a}-\mathbf{c}) = 3\mathbf{e}$

for some vector $\mathbf{e}$ with integer coordinates. But we already know that $(\mathbf{b}-\mathbf{a})-(\mathbf{a}-\mathbf{c}) = 3\mathbf{d}$. Hence both $\mathbf{b}-\mathbf{a}$ and $\mathbf{a}-\mathbf{c}$ are integer vectors multiplied by 3. Thus s and t are divisible by 3, contradicting $r+s+t \equiv 2 \pmod 3$. It follows that one of r, s, or t is congruent to 1 modulo 3, and the others are congruent to 2. Also D must be divisible by $3rs$, $3st$, and $3rt$.

We may assume that $r \geq s \geq t$. Now $3rs \leq D = 4+r+s+t$, and so

$$(3t-1)^2 \leq (3r-1)(3s-1) = 9rs - 3r - 3s + 1 \leq 13 + 3t.$$

This inequality is false for $t \geq 2$, so $t = 1$. Therefore $r \equiv s \equiv 2 \pmod 3$ and also $(3r-1)(3s-1) \leq 16$. Together, these imply that $r = s = 2$, and so $D = 9$. Now D is not divisible by $3rs$, so we have a contradiction.

Editorial comment. The proposer discovered the following result: Let a triangle have vertices at integer lattice points $(0,0)$, (b_1, b_2), and (c_2, c_2). Let $\alpha = \gcd(b_1 - c_2, b_2 - c_2)$, $\beta = \gcd(b_2, b_2)$, and $\gamma = \gcd(c_1, c_2)$. The center of mass is a lattice point if and only if either **(i)** 3 is a divisor of all three numbers α, β, and γ; or **(ii)** 3 is a divisor of none of the three numbers α, β, and γ, but 3 is a divisor of the double area $D = b_1c_2 - b_2c_1$.

Only partial solutions were received for part **(b)**. Searches by John H. Lindsey II and by the GCHQ Problems Group found the values $i < 1000$ satisfying the condition. The two lists are the same, except that 906 appears in one list and not the other. The remaining values found were: 3, 6, 15, 18, 30, 36, 48, 51, 63, 78, 90, 108, 120, 138, 150, 156, 168, 210, 228, 270, 300, 303, 336, 360, 378, 408, 426, 438, 480, 510, 528, 531, 630, 660, 723, 738, 750, 780, 888, 930, 990, 996.

Part **(a)** also solved by J. H. Lindsey II, GCHQ Problems Group (U. K.), and the proposer.

A Surrounded Set

10608 [1997, 664]. *Proposed by Victor Zalgaller, Steklov Mathematical Institute, St. Petersberg, Russia.* Let S be a compact convex set in the plane. If l is any line of support for S, let $f(l)$ be the length of the shortest curve that begins and ends on l and that together with l surrounds S. Prove that if $f(l)$ is independent of l, then S is a circle.

Solution by John Arkinstall, Monash University, Australia. Let l' be the support line parallel to l on the opposite side of S. Since $f(l) + f(l')$ is independent of l but also equals the perimeter of S plus twice the width of S perpendicular to l, S is a set of constant width w. A line l'' parallel to l and l' whose intersection with S is of maximum length is called a *diameter* of S in the direction of l. Because S has constant width w, we may conclude that the length of such a diameter is w, that it joins two points where support lines perpendicular to l touch S, that these two support lines, together with l and l', form a square of side w, and that each support line touches S in a unique point. A theorem of Khassa (Relation between maximal chords and symmetry for convex sets, *J. London Math. Soc.* **15** (1977) 541–546) states that a convex curve of constant width in which the diameter in each direction is midway between the two support lines in that direction must be a circle. Thus it suffices to prove this "midway" property.

When S is a set in the plane of constant width w, $2A(l) - wP(l)$ is independent of l, where $A(l)$ is the area of the portion of S between the opposite support line l' and the diameter l'', and $P(l)$ is the length of that part of the perimeter of S on the same side of l'' as l' (L. Beretta & A. Maxia, "Insiemi convessi e orbiformi," *Univ. Roma e Ist. Naz. Alta Mat. Rend. Mat.* (5) **1** (1940) 1–64). Let $r(l)$ denote the distance from l to the diameter in the direction of l. Since $r(l)$ is the length of the supporting line segment from l to the boundary of S on the shortest curve from l surrounding S, we have $f(l) = 2r(l) + P(l)$. Since this is independent of l by hypothesis, so is $A(l) + wr(l)$. This is the area of the convex hull of S and the two supporting line segments from l on the shortest curve from l. The complement R of this convex hull in the supporting square to S with side length w and edge along l therefore also has area independent of the direction of l.

Now consider how the area of R changes when the direction of l is changed by a small angle ϕ. If the point of support on l' divides its side of the supporting square in the ratio $u : w - u$, then the area of R changes by four small approximately triangular regions: It decreases by $(1/2)r(l')\phi + o(\phi)$, increases by $(1/2)u\phi + o(\phi)$, decreases by $(1/2)(w - u)\phi + o(\phi)$, and increases by $(1/2)r(l')\phi + o(\phi)$. Thus, the area of R changes by the sum $(1/2)(2u - w)\phi + o(\phi)$. Since this is 0, we have $2u - w = 0$, and thus the support point on l' is midway between the two support lines perpendicular to l'.

Solved also by S. S. Kim (Korea), J. G. Merickel, GCHQ Problems Group (U. K.), and the proposer.

Tight Bounds for the Normal Distribution

10611 [1997, 665]. *Proposed by Zoltán Sasvári, Technical University of Dresden, Dresden, Germany.* Find the largest value of a and the smallest value of b for which the inequalities

$$\frac{1+\sqrt{1-e^{-ax^2}}}{2} < \Phi(x) < \frac{1+\sqrt{1-e^{-bx^2}}}{2}$$

hold for all $x > 0$, where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\,dy$.

Solution by Hongwei Chen, Christopher Newport University, Newport News, VA. We show that $a = 1/2$ and $b = 2/\pi$ are the best possible constants for which the stated inequalities hold. Since $\int_{-\infty}^{0} e^{-y^2/2}\,dy = \int_{0}^{\infty} e^{-y^2/2}\,dy = \sqrt{\pi/2}$, the stated inequalities are equivalent to

$$\frac{\sqrt{1-e^{-ax^2}}}{2} < f(x) < \frac{\sqrt{1-e^{-bx^2}}}{2}, \tag{1}$$

where $f(x) = \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-y^2/2}\,dy$. If the second inequality of (1) holds for all $x > 0$, then

$$0 < \frac{\sqrt{1-e^{-bx^2}}}{2} - f(x) = \left(\frac{\sqrt{b}}{2} - \frac{1}{\sqrt{2\pi}}\right)x + O(x^3)$$

as $x \to 0$, which implies $b \geq 2/\pi$. Similarly, if the first inequality of (1) holds for all $x > 0$, then

$$0 < f(x) - \frac{\sqrt{1-e^{-ax^2}}}{2} = \frac{1}{4}e^{-ax^2} + O\left(e^{-2ax^2}\right) - \frac{e^{-x^2/2}}{\sqrt{2\pi}x} + O\left(\frac{e^{-x^2/2}}{x^2}\right)$$

as $x \to \infty$. Dividing each side by $e^{-x^2/2}$ yields $a \leq 1/2$.

To show that inequalities (1) hold for all $x > 0$ when $a = 1/2$ and $b = 2/\pi$, we write

$$(f(x))^2 = \frac{1}{2\pi} \int_0^x \int_0^x e^{-(y^2+z^2)/2}\,dydz.$$

Let $D = [0, x]^2$, $D_1 = \{(y, z) : 0 \leq y,\ 0 \leq z,\ y^2 + z^2 \leq x^2\}$, and $D_2 = \{(y, z) : 0 \leq y,\ 0 \leq z,\ y^2 + z^2 \leq (4/\pi)x^2\}$. We have the inequalities

$$\frac{1}{2\pi}\iint_{D_1} e^{-(y^2+z^2)/2}\,dydz < \frac{1}{2\pi}\iint_{D} e^{-(y^2+z^2)/2}\,dydz < \frac{1}{2\pi}\iint_{D_2} e^{-(y^2+z^2)/2}\,dydz, \tag{2}$$

the first because $D_1 \subset D$, and the second because D and D_2 have the same area and $e^{-(y^2+z^2)/2} \leq e^{-(2/\pi)x^2}$ for $(y, z) \in D - D_2$ while $e^{-(y^2+z^2)/2} \geq e^{-(2/\pi)x^2}$ for $(y, z) \in D_2 - D$. Evaluating the outer integrals in (2) in polar coordinates, we obtain

$$\frac{1-e^{-x^2/2}}{4} < f(x)^2 < \frac{1-e^{-2x^2/\pi}}{4},$$

which is equivalent to (1).

Solved also by P. Alsholm (Denmark), J. Anglesio (France), P. Bracken (Canada), B. Burdick, G. G. Chappell, P. Devaraj (India), G. Keselman, K.-W. Lau (Hong Kong), J. H. Lindsey II, A. Stadler, GCHQ Problems Group (U. K.), NCCU Problems Group, NSA Problems Group, WMC Problems Group, and the proposer.

REVIEWS

Edited by **Harold P. Boas**
Mathematics Department, Texas A & M University, College Station, TX 77843

Mathematics: From the Birth of Numbers. By Jan Gullberg, W. W. Norton, 1997, xxiii + 1093 pp., $50.

Reviewed by **Arnold Allen**

Gullberg's book has worthy competitors for the title *The People's Guide to Mathematics*. In fact, at this time there is a veritable cornucopia of excellent mathematics books for the general reader. Ones that I particularly like are those by Courant and Robbins (revised by Stewart) [**3**], Devlin [**4**, **5**], Dunham [**6**], Hildebrandt and Tromba [**8**], Jacobs [**9**], and Stewart [**12**, **13**] for surveys of mathematics. For enlightenment in more specialized areas, I like Casti [**1**], Conway and Guy [**2**], Körner [**10**], and Vilenkin [**14**]. Gullberg's book is clearly the overall winner. This book will appeal to a range of MONTHLY readers as well, from undergraduate math majors to instructors. It is a wonderful read. I take it with me everywhere I go: to the dentist's office, waiting in the doctor's office, to boring meetings—wherever I have an opportunity to read.

The book lived up to its description on the dust jacket, which included:

> This gently guided, profusely illustrated Grand Tour of the world of mathematics takes the reader on a long and fascinating journey—from the dual invention of numbers and language, through the primary realms of arithmetic, algebra, geometry, trigonometry, and calculus, to the final destination of differential equations, with excursions into symbolic logic, set theory, topology, fractals, probability, and assorted other mathematical byways. ... The text is interspersed with more than 1,000 original, high-quality technical illustrations, a multitude of reproductions from mathematical classics and other relevant works, and a generous sprinkling of humorous asides ranging from limericks and tall stories to cartoons and decorative drawings.

The sentence following the ellipsis explains some of the special charm of the book. Gullberg has made even more innovative use of the margin than Graham, Knuth, and Patashnik [**7**] did with their irreverent student "graffiti." Gullberg has excellent quotes and quips throughout the book, including the margin, but he also uses the margin for noting things mentioned in the preceding paragraph as well as for pointers to related topics, biographies of famous mathematicians, labels concerning the nearby text, photographs, and other uses too numerous to catalog.

Let me describe my ramblings in this remarkable book. The first chapter discusses the birth of numbers or "Where did numbers come from, anyway, and what are their properties?" This chapter is followed by two excellent chapters that describe systems of enumeration and types of numbers. The next chapter, "Cornerstones of Mathematics," is my favorite chapter and will appeal to beginners and old hands as well. I was amazed to learn in this chapter that the pencil-and-paper multiplication algorithm we learned in grade school is not the *only* such multiplication algorithm. Gullberg demonstrates "the Egyptian method of duplation" algorithm for multiplication. (He also displays a multiplication table

from the *Bomberger Rechenbuch* of 1483 which indicates that $7 \times 3 = 12$. Murphy's law had already been passed!) I loved his example of the "Russian peasant" method of "multiplication by successive duplation and mediation." I learned a dividing and averaging method, originating with the ancient Babylonians, for finding the square root of a number; it is efficient, too. I learned how to approximate $\sqrt{6}$ by a continued fraction and how to convert 221/41 to the continued fraction

$$5 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{3 + \cfrac{1}{2}}}}}$$

I saw Pascal's triangle as Pascal drew it. Gullberg shows it as found in Pascal's book *Traité du Triangle Arithmétique*, Paris, 1665. We see it in another form from a book written in 1303 by the Chinese mathematician Chu Shih-chieh. Pascal was unaware of this earlier work and drew his triangle very differently from the way we now draw it. I learned how to use an abacus for multiplication and division. The decorative drawings in Figure 1 appear in the margin of the book next to the last four instructions for dividing 377 by 26 with the abacus.

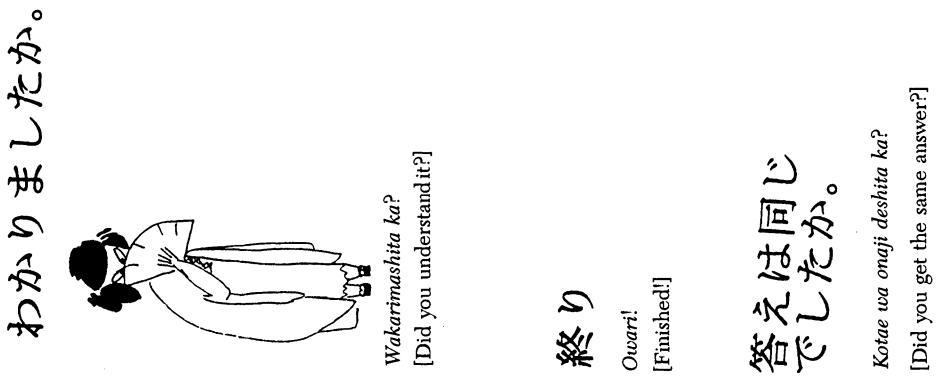


Figure 1

I was delighted by Gullberg's clear and brief chapter on combinatorics. In it he discusses several interesting combinatorial problems and their solutions. I especially liked the section on graph theory and the section on magic squares and their kin. In the former, he discusses the famous seven bridges of Königsberg problem. He even provides a map of Königsberg circa 1740! He describes Euler's solution of the problem as well as the history of the related four-color map problem. Gullberg describes and displays several famous magic squares, such as Dürer's remarkable order 4 magic square, constructed in 1514, and Benjamin Franklin's order 8 magic square with some unbelievable properties. Yes, *the* Benjamin Franklin. Franklin's magic square shown in Figure 2 appears in the margin of Gullberg's book.

Fibonacci's sequence and related sequences are well treated in the chapter on sequences and series. Immediately after his statement of Fibonacci's original

FRANKLIN Benjamin
(1706–1790)

52	61	4	13	20	29	36	45
14	3	62	51	46	35	30	19
53	60	5	12	21	28	37	44
11	6	59	54	43	38	27	22
55	58	7	10	23	26	39	42
9	8	57	56	41	40	25	24
50	63	2	15	18	31	34	47
16	1	64	49	48	33	32	17

Franklin's Magic Square

Figure 2

problem, Gullberg provides an amusing drawing of a cluster of rabbits. Then, in a magical two pages, he provides a great deal of interesting information about the Fibonacci (F_n), Lucas (L_n), and Pell (P_n) sequences, including a discussion of the golden ratio and the Binet formulas for F_n and L_n. Another peak experience for me in this chapter was reading the section on figurate numbers. Using graphic software to produce displays representing numbers by dots, Gullberg develops the properties of triangular, oblong, and pentagonal numbers in the plane. Then, using three-dimensional displays of dots, he derives the properties of cubic, tetrahedral, and square pyramidal numbers. Finally, he demonstrates how to extend tetrahedral numbers to superhedral numbers. An enjoyable chapter!

Fermat's last theorem has been in the news, but I found it difficult to explain to friends. Now I refer them to Gullberg's two-page discussion; it is at exactly the right level for those who are mathematically disadvantaged, but with some sophistication as well. Friends do not have to endure a discussion of elliptic curves and the Taniyama-Shimura conjecture as a journalism-educated friend of mine did when she bravely attended a math meeting with me.

The chapter "Theory of Equations" that contains the material on Fermat's last theorem also demonstrates, with worked examples, how to solve almost any kind of equation: quadratic, cubic, quartic, Diophantine; systems of linear equations, systems of linear and quadratic equations; and others. A pleasure to read!

I enjoyed Gullberg's brief but fascinating "Overture to the Geometries." It has the history and a quick sketch of the various branches. There are nice pictures, illustrations, and diagrams, including some elegant drawings of knots and other topological entities. I also liked the five chapters on geometry that follow, especially the section in the trigonometry chapter on solving triangles. Gullberg shows how to keep difficult geometry problems from becoming boring; he solves them with élan.

The most complete coverage of any subject in the book is given to mathematical analysis. There are chapters on differential and integral calculus and power series; detailed chapters on the major applications of calculus; a chapter on harmonic analysis with Fourier series; a chapter on methods of approximation; and an excellent introductory chapter on differential equations. Each chapter features theory as well as the solution of non-trivial problems.

I loved the probability chapter. There is a wonderful section on the history of probability with the names and achievements of most of the major players.

Gullberg also provides an excellent survey, with examples, of basic probability and statistics. He describes the basic distributions, such as normal and binomial, with their distribution functions and examples of how they are used. There is a photograph of a mustached gambler and his moll that looks suspiciously like Gullberg and his wife.

I bought my copy of Gullberg's book after leafing through it at a bookstore. It impressed me so much that I showed it to my colleagues at the office (I am part of a small R & D group). Everyone wanted to borrow it! I had to hide it in my desk to keep it from disappearing. One colleague got his wife a copy for her birthday—she is a high school math teacher. She loved it and showed it to their son's math teacher who is now an enthusiastic owner, too. To improve mathematics education in the USA we might begin by providing every high school math teacher with a copy of the book.

Later I learned that the large first printing was sold out almost immediately. Then I discovered that the Book-of-The-Month-Club had chosen it as one of its selections, and that the MAA has made it available to its members. The book has had the success that an author might dream of on writing a first book. Sometimes there is a bit of luck involved in becoming a best-selling author, perhaps an invitation to be a guest on Oprah Winfrey's show. But the success of Gullberg's book is not due to luck, or to Oprah: this really is a wonderful book! The dust jacket contains rave reviews from Martin Gardner, Philip Morrison, and Harold Jacobs. In addition, Peter Hilton wrote a forward "Mathematics in Our Culture."

What is the appeal of this book to the general reader? I could use a real estate analogy and say "Coverage, coverage, coverage." The book has excellent chapters on the standard material that most MONTHLY readers—as well as many engineers, scientists, and general readers (the "two cultures folks")—have studied somewhere before but would like to review. In addition, there are special chapters or sections of chapters on subjects that many of us have never seen before—such as the first several chapters that cover the birth of numbers and their applications; two algorithms in Chapter 4 for finding cube roots (Gullberg says, "We have found the good physician Trenchant's method an excellent cure for insomnia") as well as several methods for approximating cube roots; the section on the abacus, the slide rule, Napier's bones, and the quipu; and the incredible Chapter 11, "Overture to the Geometries," with much interesting material known only to specialists.

I am surrounded by engineers who do not use mathematics on a daily basis but who occasionally have special problems that need some mathematical analysis. Since I am the only one around who has a Ph.D. in mathematics, they sometimes come to me to ask basic mathematical questions. Now I can just refer them to the Gullberg book for standard material. I believe it will become a general reference for many MONTHLY readers. There are other uses of the book for MONTHLY readers. The possibilities for undergraduate student readers include:

- Reading the book to broaden their understanding of mathematics beyond their course work.
- Using it as a supplement to many of their courses, such as calculus, college algebra, geometry, and statistics.

The possibilities for instructors include the following:

- Using it as a textbook for courses such as *Mathematics for Poets* or *Calculus for Poets*. An experiment like this is being tried by Dennis Watson and his colleagues at Clark College in Vancouver, Washington.

- Using it as a supplementary textbook.
- Using it for extra credit research study and reports.

As an author, I must note possible uses of the book for authors of undergraduate textbooks. These include:

- Saving time and money for special figures or material by referring their readers to the book, rather than including the special material in their books. For example, I plan to refer readers of the book I'm writing to Gullberg's book for the history of probability and to show them what Pascal's triangle looked like to Pascal as well as to ancient Chinese mathematicians. This will be better than writing a history myself or getting the necessary permissions to make copies of the Pascal triangles.
- Using it as a source of ideas for material to discuss and for examples to emulate.

Gullberg's book is a gigantic book in every sense of the word. It weighs four pounds and thirteen ounces and has 1100 pages. (While reading this book is most enjoyable, it gives a whole new meaning to the phrase "heavy reading.") It took a giant effort to produce such a masterpiece. Gullberg spent more than ten years writing it, mostly at night after a full day as an active physician and surgeon. However, he got a lot of help, especially from his family. His son persuaded him to write it, provided most of the mathematical graphics, helped him keep his computer running, and counseled him about content. His wife, Ann, proofread the manuscript and drew many of the illustrations. Some others were done by their twin sons, Kamen and Kalin, when they were nine or ten years old. Gullberg also has talented friends in various disciplines who reviewed what he was doing, and he got some help from outside professionals. As he says in the preface, "My draft text was later polished with the help of distinguished professional mathematicians, linguists, and historians."

Although he got others to help him with the writing of the book, he did the actual writing using a Macintosh Plus and Microsoft Word 3.0. These were state-of-the-art when he began his writing, but are ancient relics now. It was a most challenging task to keep the Macintosh Plus going all those years and to hack the equations without LaTeX or a commercial software package. However, he did it. As he notes on the copyright page: "Camera-ready copy for this book was produced entirely by the author, utilizing a combination of modern desktop-publishing and traditional paste-up methods." This was an incredible achievement.

Peter Renz [**11**] describes some of the problems Gullberg had with his computer and how and why the book was written. Renz also gives a biographical sketch of Gullberg and some details of the production and publication of the book.

These are my three wishes for an encore from Dr. Jan Gullberg:

- Set up a web site so that we can have two-way communication with him. For example, instructors who are using his book could let him know why they decided to use it, and Dr. Gullberg could make this information available to others. Everyone could suggest new exercises, or potential improvements, or corrections, and he could post additional material, corrections, etc.
- Write a solutions manual for the exercises or commission someone to do so.
- Write a second volume containing material he hadn't time or room enough to put in the present book, such as Green's theorem, wavelets, encryption techniques such as trap door ciphers—the kind of material that is described

in Barry Cipra's series for the American Mathematical Society on what's happening in the mathematical sciences.

Jan Gullberg's book is a giant leap forward for mathematics and all those who love it!

REFERENCES

1. John L. Casti, *Five Golden Rules*, Wiley, 1996.
2. John H. Conway and Richard K. Guy, *The Book of Numbers*, Springer, 1996.
3. Richard Courant and Herbert Robbins, *What is Mathematics?*, second edition revised by Ian Stewart, Oxford University Press, 1996.
4. Keith J. Devlin, *Mathematics: The Science of Patterns*, W. H. Freeman, 1994.
5. Keith J. Devlin, *Mathematics: The New Golden Age*, Penguin Books, 1988.
6. William Dunham, *The Mathematical Universe*, Wiley, 1994.
7. Ronald L. Graham, Donald E. Knuth, and Oren Patashnik, *Concrete Mathematics: A Foundation for Computer Science*, 2nd ed., Addison-Wesley, 1994.
8. Stefan Hildebrandt and Anthony Tromba, *The Parsimonious Universe*, Springer, 1996.
9. Konrad Jacobs, *Invitation to Mathematics*, Princeton University Press, 1992.
10. T. W. Körner, *The Pleasures of Counting*, Cambridge University Press, 1996.
11. Peter Renz, A Long Loving Look at Mathematics, *Math Horizons*, April 1997, 16–22.
12. Ian Stewart, *Nature's Numbers*, Basic Books, 1995.
13. Ian Stewart, *From Here to Infinity: A Guide to Today's Mathematics*, Oxford University Press, 1996.
14. Naum Yakovlevich Vilenkin, *In Search of Infinity*, Birkhäuser Boston, 1995.

Hewlett-Packard Advanced Technology Center, 8000 Foothills Boulevard, MS 5740, Roseville, California 95747
arnolda@jps.net

EDITOR'S CORNER. I am a mathematician today because of an experience in kindergarten. I have always been interested in analyzing patterns, and from an early age I was excited by all sorts of intellectual games having a strong element of pattern recognition: chess, go, Scrabble, crossword puzzles, and so forth. The first really sophisticated pattern-recognition task that young children face is learning to read, and I was starting to read at home before I began formal schooling. However, because reading was on the syllabus for first grade, my kindergarten teacher viewed reading as a proscribed activity and successfully quashed my initial enthusiasm for the printed word. Consequently, I redirected my energies toward arithmetic, which the teacher apparently did not recognize as a subversive subject. The result was that later on, I was always a year ahead of my schoolmates in mathematics, and so I ended up a mathematics major in college. Thus, it is thanks to my kindergarten teacher that I am today a specialist on the theory of the Bergman kernel function rather than on, say, the works of Jane Austen.

There is a moral in this story for the public school system, but I should like to address instead what this story suggests about how to teach mathematics effectively. It is a familiar metaphor that mathematics is a language, and in my early experience mathematics and language were both difficult—and therefore interesting—for the same reason. Both subjects to me were games presenting an artificial universe operating by well defined, albeit arbitrary, rules. The goal was to understand the rules and their consequences so thoroughly that they became internalized and automatic, in which case one could justifiably claim to have solved the game. It is in this sense that a sixth-grader who can unscramble Rubik's cube rapidly while blindfolded can assert mastery of that game.

Now if mathematics is a language, why are the standard methods of teaching mathematics not even remotely similar to the standard methods of teaching

languages? The way to become fluent in a language is total immersion. Has anyone tried to foster mathematical ability by the same method? Actually, yes: intensive summer programs for teenagers, such as the famous one developed at Ohio State by Arnold Ross, have indeed proved effective. However, the standard mathematics program in a typical American school is more akin to wading than to immersion. If our children spend one-fourth the time on mathematics that they spend watching television, is it any wonder that they fared poorly in the recent Third International Mathematics and Science Study? (See `http://wwwcsteep.bc.edu/timss` for detailed information about TIMSS.)

Children seem automatically to soak up their native language. A foreign language is a better comparison to mathematics. In French class, for example, students get practice in speaking French, reading French, and writing French; drill on French vocabulary and grammar; and exposure to French culture and history. Compare this with mathematics instruction. It is an uncommon mathematics class that pays much attention to mathematical culture and history. Although the so-called "reform" movement supports giving students practice on reading and writing mathematics, this idea is spreading slowly. If you have not tried the experiment, assigning your mathematics class a writing task will be an eye-opening experience. At the beginning of the semester, I assigned a class of engineering students to write a short essay interpreting part of the introductory chapter in their differential equations textbook. Ten percent of the students immediately dropped the course. I was chagrined to discover that of those students who completed the assignment, about one-third had not understood the book's discussion of implicit solutions versus explicit solutions because they did not know the meaning of the word "implicit." It is a struggle to teach students in a way that is different from the way they were taught in grade school and high school.

Please note that I am not complaining that students are getting worse and that the world is going to the dogs. (This is true, and has been true throughout recorded history, but that is another story.) Actually, teaching college mathematics is a rewarding experience if one takes advantage of the skills that modern students have. My students cannot do arithmetic rapidly and may not comprehend words derived from Latin roots, but they have well developed social skills, know how to work collaboratively in groups, can give coherent oral presentations, and are highly creative if given a little encouragement. I fondly remember one response to an examination question I posed requiring the students to invent a scenario involving a swimming pool for which a given differential equation would be an appropriate model. The solution I expected was a mixing problem concerning the concentration of chlorine, but one student's (mostly correct) answer began: "Kim drops a calculator into a swimming pool from a height of 100 meters . . . ." It is not the case that today's students cannot think, but it is true that their intellectual and cultural backgrounds are different from those of their teachers. My point is that, like generals who prepare to fight the previous war, textbook expositors tend to write prose for the previous generation.

In a standard mathematics book, what gets taught is primarily mathematical vocabulary and rules of mathematical grammar. Perhaps in advanced undergraduate classes, students may begin to learn some principles of mathematical composition, but it is only in graduate school that one gets exposed to the mathematical analogues of fine literature and poetry.

It would be wonderful to have some textbooks that go beyond mathematical vocabulary and grammar to incorporate the mathematical equivalents of literature, poetry, culture, and history in a way that students can understand and appreciate.

Reading Arnold Allen's review of Jan Gullberg's *Mathematics: From the Birth of Numbers* made me think that it might be such a book, so I obtained a copy and read it. My impressions of the book are those of a university professor of mathematics active in both research and teaching.

The first impression one gets from Gullberg's massive tome is that producing it was a tremendous effort. Gullberg is a remarkable person: not a professional mathematician, he had the motivation, enthusiasm, and dedication to spend a decade learning about mathematics and recording his findings. His book is an epic written by an amateur.

It is not surprising that the best parts of the book are the ones that do not involve advanced mathematics. An amateur has the opportunity for originality of presentation when discussing the history of mathematics and those topics generally known as "recreational mathematics." It is here that Gullberg shines. Even professionals will find some amusing tidbits in the book. For example, I did not know that the circumcenter, centroid, and orthocenter of a triangle are on the same line, and reading this statement in Chapter 12 sent me to the library to look for a geometry book with proofs in it.

The second half of the book, where Gullberg tackles calculus and other parts of the university mathematics curriculum, disappointed me. Presumably Gullberg learned his calculus from a standard textbook, and his sections on calculus read like a distillate of any traditional calculus book. In Gullberg's encyclopedic work, there is little space for discursive exposition, and the impression that comes across is that mathematics consists of rules, formulas, and algorithms. If an educated layperson with extraordinary motivation has this concept of mathematics after studying our textbooks, then we writers of mathematics books have failed miserably.

In any book of over 2^{10} pages, there are mistakes. I found fewer than expected. There are some typographical errors, such as on page 570 where the eccentricity of an ellipse is given as $\epsilon < 0$ instead of $\epsilon < 1$. Then there are some mathematical misconceptions, such as the confused statement and proof of Cauchy's mean-value theorem on pages 718–719, where Gullberg demands that the two functions have matching values at the endpoints (which makes the theorem rather trivial). I found only one true howler: on page 735, Gullberg comes unstuck calculating the indefinite integral of the function f defined by $f(x) = 1/(\sqrt[3]{x} + \sqrt[4]{x})$ and obtains a wrong answer that is even singular at $x = 1$.

Gullberg is writing at the limit of his knowledge, and it shows. Every college mathematics teacher who reads this book will wince at some statements that are not exactly wrong, but that will certainly mislead naive readers. Here are some examples—admittedly taken out of context—of statements that make me queasy.

- "with the assistance of computers, we can easily obtain approximate values of virtually any convergent series" (page 282)
- "points on the graph of a function may approach successively nearer to a straight line, the **asymptote**" (page 341)
- "**topology** is a **non-metric geometry**" (page 369)
- "Unlike matrices, which may be of any rectangular shape, determinants are always square." (page 646)
- "$i^i = e^{-\pi/2}$" (page 793)
- "A differentiable function of one independent variable has a maximum or a minimum where its first derivative is zero." (page 816)

Exercise. Why do I object to each of these statements?

Many readers will find Gullberg's book valuable, despite its imperfections. However, it is not the book I hoped for. Although it has history and culture, and lots of vocabulary and grammar, it does not get to literature and poetry; nor does it give any inkling that most of the mathematics that exists was created in the twentieth century. I would not dare to give this book to my niece, for it would only confirm her belief that mathematics is just boring formulas, but I would have been a grateful recipient if someone had given me this book when I was a high school freshman. It would have opened my eyes to what an extensive game mathematics is, and I would have had fun playing with the formulas and trying to figure out why they work.

Gullberg produced camera-ready copy for this book himself, which was a tremendous job. I wonder why he did not enlist the aid of a professional typesetter, who could have warned him against solecisms such as increasing the inter-letter spacing of a widow word on the last line of a paragraph. The typesetting is a metaphor for the mathematics in the book: it is serviceable, but it does not look quite right to an expert eye.

I continue to wait for someone to write a book that treats mathematics as a foreign language. I want to give a copy to my kindergarten teacher.

—Harold P. Boas

Handbook of Applied Cryptography. By Alfred J. Menezes, Paul C. van Oorschot, and Scott A. Vanstone. CRC Press, Boca Raton, 1997, 780 + xxviii, $79.95.
The Cryptographic Imagination: Secret Writing from Edgar Poe to the Internet. By Shawn James Rosenheim. Johns Hopkins University Press, Baltimore, 1997, 264, $47.50.

Reviewed by **Jeffrey Shallit**

Cryptography is the art and science of writing and reading concealed messages. David Kahn, in his monumental work, *The Codebreakers* [**9**], traces its origin to 1900 B.C.E., in an inscription carved on the tomb of Khnumhotep II. Not surprisingly, modern cryptography has a strong mathematical component. Indeed, just weeks before the Japanese attack on Pearl Harbor in World War II—a war where cryptography would play a crucial role—algebraist A. A. Albert stated in a regional meeting of the American Mathematical Society, "It would not be an exaggeration to state that abstract cryptography is identical with abstract mathematics." [**1**]

As we make the transition to a wired society, cryptography is becoming more and more fundamental. Cryptography can be used to preserve the privacy of messages exchanged over the Internet, to provide a means for electronically signing messages (so that recipients can be assured the message really came from the stated author), to prevent unauthorized access to sensitive information, and to provide a secure system for "electronic cash." Of course, like any useful tool, it can also be used to conceal evidence of lawbreaking, which is one reason why the debate over cryptography policy is so heated [**8**]. The United States government continues to restrict export of certain kinds of cryptographic software, although

lately there have been some signs of weakening of this policy. As I write this, Senators Conrad Burns (R MT) and Patrick Leahy (D VT) have introduced the "Promotion of Commerce Online in the Digital Era (Pro-CODE) Act," a bill designed to relax government controls on encryption.

Although some (notably Dorothy Denning [**5**]) continue to endorse some kinds of government control over cryptography, most observers concede that the genie is out of the bottle, and no law can stuff it back in. Some of the mathematics involved in cryptography is so elementary that a bright high-school student can devise encryption software essentially uncrackable with currently available techniques.

Modern cryptography has traveled a long way from the simple substitution cipher, where messages are encoded by applying a permutation of the letters $\{A, B, C, \ldots, Z\}$. As a glance at the proceedings of the latest conferences on the subject (CRYPTO, EUROCRYPT, AUSCRYPT, ASIACRYPT) demonstrates, modern cryptography is based largely on number theory, combinatorics, and algebra. Although some cryptographic techniques are quite elementary, others involve the deepest and most modern aspects of these areas.

The so-called *public-key cryptosystem* is a fundamental discovery of modern cryptography. In older methods based on secret keys, the same key is used to encode and decode a message. Losing the key means that the secrecy of the message is compromised. In 1976, however, Whit Diffie and Martin Hellman invented a new kind of cryptosystem in which the encryption and decryption keys differ [**6**]. To employ the system, a user U publishes the encryption key e for all to see. By using the key, anyone can send an encoded message to U. But U keeps the decryption key d a secret, and the system is designed so that obtaining d from e is a problem that appears to be computationally intractable. In practice, then, without enormous computing resources or extraordinary luck, only U can read the message because only U knows the decryption key d.

The RSA scheme—the name comes from the initials of its inventors, Rivest, Shamir, and Adleman—is the most famous example of a public-key cryptosystem [**11**]. It is based on the apparent intractability of computing kth roots modulo composite numbers. Invented in 1977, the system continues to withstand the various attacks devised against it, although research has provided some caveats about how best to choose the particular parameters involved. More recent proposals are based on number theory, and draw their inspiration from the discrete logarithm problem, the quadratic residuacity problem, and the theory of elliptic curves.

Another advance of modern cryptography is the *digital signature*. Here the goal is to attach a number to a message depending on (a) some secret key known only to the signer and (b) the content of the message itself. A good digital signature scheme allows the user to verify a message's authenticity and prevents later repudiation of signatures. Secure digital signatures are essential if electronic commerce is to become a reality.

In their *Handbook of Applied Cryptography*, Menezes, van Oorschot, and Vanstone convincingly demonstrate the serious mathematical content of applied cryptography. The authors have significant expertise in both theoretical and practical aspects of cryptography. Vanstone, for example, is a founder of a corporation that makes encryption products, and is the author of at least two dozen papers in the literature on cryptographic topics. Menezes wrote his Ph.D. dissertation on public-key cryptography [**10**], and van Oorschot is employed as a cryptographer.

The approach taken in this book is encyclopedic. Nearly every aspect of modern cryptography—stream ciphers, block ciphers, public-key cryptography, hash func-

tions, digital signatures, key establishment protocols, etc.—is covered in its 800 pages. Each chapter contains a "Notes" section, with valuable historical remarks. The book contains over 1200 citations to the literature.

The *Handbook* strongly emphasizes practical aspects of cryptography. For example, Chapter 14 discusses how to perform extended-precision arithmetic efficiently (essential for modern cryptography, which routinely deals with numbers of 100 to 400 digits).

The *Handbook* probably isn't appropriate as a textbook or gentle introduction to cryptography—for that, see instead [**3**] or [**12**]. It is probably more appropriate as a reference work, and in that it succeeds admirably. There are, however, some minor flaws. For example, the surname of Franz Mertens is misspelled twice. More seriously, Definition 2.73 erroneously states that "A problem is **NP**-hard if the existence of a polynomial-time algorithm for its solution implies that **P** = **NP**." While this is correct if, as most suspect, **P** ≠ **NP**, it is incorrect if **P** = **NP**. The correct definition is that a problem X is **NP**-hard if every problem in **NP** reduces to X in polynomial time. Similarly, Definition 3.1, which defines polynomial-time reduction, is regrettably imprecise, drawing no distinction between Turing reduction and many-one reduction. Other errors can be found in the world-wide-web page for the book, `http://www.dms.auburn.edu/hac/`.

Another annoyance is that the index is inadequate. For example, there is no entry for either "**NP**" or "nondeterministic polynomial time." Also missing is an index to names, so that it is very difficult to determine where a particular paper is cited in the text. Nevertheless, despite these minor problems, the *Handbook* should prove to be a very useful reference for mathematicians and cryptographers.

The contrast between the *Handbook* and Shawn Rosenheim's *The Cryptographic Imagination* couldn't be more marked. Rosenheim, a professor of English and American Studies at Williams College, has produced a rambling, disjointed meditation on cryptography, Edgar Allan Poe, espionage, Thomas Pynchon, and the Internet.

Unfortunately, Rosenheim's attempts to discuss technical matters are nearly always marked by severe misunderstandings of the mathematics and physics involved. For example, consider his definition of quantum cryptography, which appears in the glossary:

> A form of cryptography in which, under certain experimental conditions, pairs of photons may be created that exert an influence over one another that cannot be explained by quantum mechanics. Measuring the polarization of one particle immediately and identically changes the spin on its antiparticle. Such polarization takes place regardless of the relative positions of the two particles in the universe, in a result that seems to violate the second law of classical theory. It is theoretically possible that a stream of such polarized photons could be used to encipher messages that could be sent over space in literally no time at all.

There are so many errors in just these four sentences that it is difficult to know where to begin. First of all, the behavior of entangled photon pairs is, contrary to the claim, perfectly explicable through quantum mechanics. Second, practical quantum cryptography is not currently based on entangled photon pairs—although Ekert [**7**] did propose such a scheme—but a different mechanism proposed much earlier by Wiesner [**13**] and Bennett and Brassard [**3**]. Third, the reference to the "second law" is, of course, utter nonsense. Finally, quantum cryptography is not simply a theoretical possibility, but a practical reality [**2**].

Other blunders in *The Cryptographic Imagination* include conflating monkeys and apes, misstating Zipf's law, wildly overestimating the amount of pornography

on the Internet, misstating the name of the Usenet newsgroup alt.sexual.abuse.recovery, and comically misspelling the name of one of the inventors of RSA as "Ronald Rivers." Rosenheim even makes mistakes in his own field: he claims that Georges Perec's book *La Vie: Mode d'Emploi* was written without the letter "e," when in fact it is another book by Perec entitled *La Disparition.*

This is not to say that I didn't get anything out of Rosenheim's book. I was intrigued to learn about Lizzie Doten, a 19th century mystic who "channeled" ersatz poems of Poe and other writers such as Shakespeare and Burns. But the book is marred by the usual postmodernist excesses: making much of tenuous or nonexistent connections, second-rate wordplay (the series in which *The Cryptographic Imagination* is published is entitled "re-visions of culture and society"; among postmodernists, this sort of gratuitous hyphen insertion is apparently considered essential), and opaque exposition. Consider the following two examples:

> When I claim that Poe helped end World War II, the "Poe" in that sentence represents both a particular author and the literary genre he helped create and for which he serves as a synecdoche. (p. 15)

> Such a homeopathic technique for the creation of mysteries produces highly cathected readers; the surface of the cipher produces a crypt in us, which we proceed to fill with our imagination, just as the semantic vacuity of Khumnhotep's [sic] glyphs contextually signified Khumnhotep's [sic] power and his resistance to comprehension. (p. 48)

The Cryptographic Imagination will be of little interest to anyone wanting to learn about cryptography. In fact, I can scarcely think of a reason to read it, except perhaps to see an example of what passes for scholarly work in some academic disciplines.

REFERENCES

1. A. A. Albert, Some mathematical aspects of cryptography, in R. E. Block et al., eds., *A. Adrian Albert: Collected Mathematical Papers*, American Mathematical Society, 1993, pp. 903–920.
2. C. H. Bennett, F. Bessette, G. Brassard, L. Salvail, and J. Smolin, Experimental quantum cryptography, *J. Cryptology* **5** (1992) 3–28.
3. G. Brassard, *Modern Cryptology*, Lecture Notes in Computer Science #325, Springer-Verlag, 1988.
4. C. H. Bennett and G. Brassard, Quantum cryptography and its application to provably secure key expansion, public-key distribution, and coin-tossing, *IEEE Int. Symp. Information Theory*, September 1983, p. 91.
5. Dorothy E. Denning, Resolving the encryption dilemma: the case for Clipper, *Technology Review* **98** (5) (July 1995) 48–55.
6. W. Diffie and M. E. Hellman, New directions in cryptography, *IEEE Trans. Info. Theory* **22** (1976) 644–654.
7. A. K. Ekert, Quantum cryptography based on Bell's theorem, *Phys. Rev. Lett.* **67** (1991) 661–663.
8. Lance J. Hoffman, *Building in Big Brother: The Cryptographic Policy Debate*, Springer-Verlag, 1995.
9. David Kahn, *The Codebreakers: The Story of Secret Writing*, Macmillan, 1967.
10. A. R. Menezes, *Elliptic Curve Public Key Cryptosystems*, Kluwer, 1993.
11. R. L. Rivest, A. Shamir, and L. Adleman, A method for obtaining digital signatures and public-key cryptosystems, *Commun. ACM* **21** (1978) 120–126.
12. D. R. Stinson, *Cryptography: Theory and Practice*, CRC Press, 1995.
13. S. Wiesner, Conjugate coding, *SIGACT News* **15** (1) (1983) 78–88.

Department of Computer Science, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada
shallit@graceland.uwaterloo.ca

TELEGRAPHIC REVIEWS

Edited by **Arnold Ostebee**
with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges

Telegraphic Reviews are designed to alert readers in a timely manner to new books appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T : Textbook	*P : Professional Reading*	*1–4: Semester*
C : Computer Software	*L : Undergraduate Library*	*** : Special Emphasis*
S : Supplementary Reading	*13: Grade Level*	*?? : Questionable*

Readers are advised that price information is subject to change. Selected books receive a second, more extensive review in the *Monthly*.

Books submitted for review should be sent to *Book Reviews Editor, American Mathematical Monthly, St. Olaf College, 1520 St. Olaf Avenue, Northfield, MN 55057-1098.*

General, P, L.** *Calendrical Calculations.* Nachum Dershowitz, Edward M. Reingold. Cambridge Univ Pr, 1997, xxi + 307 pp, $22.95 (P); $64.95. [ISBN 0-521-56474-3; 0-5621-56413-1] Algorithms for calculations in 14 calendars of current and historical interest (including the Gregorian, Julian, Islamic, Hebrew, Mayan, Chinese, and Hindu calendars). Implementations of the algorithms (in Lisp) are given in an appendix. Many interesting notes and comments on history of chronology. AO

General, P. *Algebra and Operator Theory.* Eds: Yusupdjan Khakimdjanov, Michel Goze, Shavkat A. Ayupov. Kluwer Academic 1998, viii + 250 pp, $114. [ISBN 0-7923-5094-4] Proceedings of a 1997 French–Uzbek colloquium in Tashkent, Uzbekistan.

General, P, L. *World Directory of Mathematicians 1998, 11th Edition.* International Mathematical Union. AMS, 1998, xxii + 1093 pp, $65 (P).

Reference, P, L. *Handbook of Mathematics and Computational Science.* John W. Harris, Horst Stocker. Springer-Verlag, 1998, xxviii + 1028 pp, $29.95. [ISBN 0-387-94746-9] Comprehensive collection of commonly used definitions, facts, formulae, tables, techniques, etc. Includes numerical as well as analytic techniques, and a brief introduction to several computer programming languages. AO

Education, P*. *Research in Collegiate Mathematics Education, III.* Eds: Alan H. Schoenfeld, Jim Kaput, Ed Dubinsky. CBMS Issues in Math. Educ., V. 7. AMS, 1998, xii + 313 pp, $40 (P). [ISBN 0-8218-0882-6] Papers describe methodology and research on problem solving, understanding of concepts and proofs.

History, P. *Paul Dirac: The Man and His Work.* Abraham Pais, *et al.* Cambridge Univ Pr, 1998, xv + 124 pp, $19.95. [ISBN 0-521-58382-9] Lectures about Dirac's life and work by A. Pais, M. Jacob, D.I. Olive, and M.F. Atiyah as well as a memorial address given by Stephen Hawking.

Logic, P. *Advances in Modal Logic, Volume 1.* Eds: Marcus Kracht, *et al.* CSLI Lect. Notes, No. 87. CSLI Pub (Center for Study of Language & Information, Leland Stanford Junior Univ., Stanford, CA 94305), 1998, xvi + 392 pp, $24.95 (P). [ISBN 1-57586-102-X] Papers from a 1996 conference at the Free University of Berlin. Thematic areas: mathematics of modal logic; artificial intelligence and knowledge representation; computation; philosophy and language; proof theory.

Logic, P. *The Tbilisi Symposium on Logic, Language and Computation: Selected Papers.* Eds: Jonathan Ginzburg, *et al.* Stud. in Logic, Lang. & Inform. CSLI Pub (Center for Study of Language & Information, Leland Stanford Junior Univ., Stanford, CA 94305), 1998, xxxi + 376 pp, $24.95 (P). [ISBN 1-57586-098-8] 23 papers from a 1995 conference in Gudauri (Republic of Georgia). Topics: natural language semantics; dynamic semantics and channel theory; theoretical linguistics; computational linguistics; formal logic theory; mathematical logic; theoretical computer science.

Number Theory, P. *The Theory of Partitions.* George E. Andrews. Math. Lib. Cambridge Univ Pr, 1998, xvi + 255 pp, $29.95 (P). [ISBN

0-521-63766-X] Republication of the 1976 Addison-Wesley edition (TR, June–July 1977).

Group Theory, P. *The Atlas of Finite Groups: Ten Years On.* Eds: Robert Curtis, Robert Wilson. London Math. Soc. Lect. Note Ser., V. 249. Cambridge Univ Pr, 1998, xvii + 293 pp, $44.95 (P). [ISBN 0-521-57587-7] 20 articles from a 1995 conference in Birmingham, England.

Group Theory, P. *Representations and Cohomology, II: Cohomology of Groups and Modules.* D.J. Benson. Stud. in Adv. Math., V. 31. Cambridge Univ Pr, 1998, xi + 279 pp, $29.95 (P); $69.95. [ISBN 0-521-63652-3; 0-521-36135-4] Paperback edition.

Ring Theory, P. *Representations of Affine Lie Algebras.* Vyacheslav M. Futorny. Papers in Pure & Appl. Math., V. 106. Queen's Univ, 1997, xi + 89 pp, (P). [ISBN 0-88911-756-X]

Topological Groups, P. *Positivity in Lie Theory: Open Problems.* Eds: Joachim Hilgert, *et al.* Expos. in Math., V. 26. Walter de Gruyter, 1998, xii + 290 pp, $138.95. [ISBN 3-11-016112-5] 15 articles provide background and pose open problems in a variety of settings.

Algebra, P. *Mixed Motives.* Marc Levine. Math. Surv. & Mono., V. 57. AMS, 1998, x + 515 pp, $109. [ISBN 0-8218-0785-4]

Calculus, T(15–16: 2), P. *A Course in Advanced Calculus.* Robert S. Borden. Dover, 1998, xii + 402 pp, $12.95 (P). [ISBN 0-486-67290-5] Unabridged, slightly corrected republication of the 1983 North Holland/Elsevier Science Publishing Co. edition (TR, March 1984). Includes a new chapter on "Tips and Solutions for Selected Problems."

Real Analysis, T(17: 4), L. *Modern Analysis.* Kenneth Kuttler. Stud. in Adv. Math. CRC Pr, 1998, 572 pp. [ISBN 0-8493-7166-X] Text for a first-year graduate course on Lebesgue integration and functional analysis. Covers standard topics (i.e., general set theory and topology, calculus in Banach spaces, Hilbert spaces, L^p spaces, representation theorems, Fourier transforms) as well as some less standard topics: probability, Bochner integral, convex functions. KS

Differential Equations, P. *Advances in Differential Equations and Mathematical Physics.* Eds: Eric Carlen, Evans M. Harrell, Michael Loss. Contemp. Math., V. 217. AMS, 1998, ix + 221 pp, $35 (P). [ISBN 0-8218-0861-3] 14 papers from the 1997 Georgia Tech–UAB International Conference on Differential Equations and Mathematical Physics.

Differential Equations, P. *Generalized Quasilinearization for Nonlinear Problems.* V. Lakshmikantham, A.S. Vatsala. Math. & Its Applic., V. 440. Kluwer Academic, 1998, ix + 276 pp, $140. [ISBN 0-7923-5038-3]

Partial Differential Equations, T(17–18). *Theory and Applications of Partial Differential Equations.* Piero Bassanini, Alan R. Elcrat. Math. Concepts & Methods in Sci. & Eng., V. 46. Plenum Pr, 1997, ix + 439 pp, $115. [ISBN 0-306-45640-0]

Partial Differential Equations, T(17–18). *Partial Differential Equations and Boundary Value Problems.* Viorel Barbu. Math. & Its Applic., V. 441. Kluwer Academic, 1998, xii + 277 pp, $129. [ISBN 0-7923-5056-1]

Numerical Analysis, S(18), P*, L*. *Computational Integration.* Arnold R. Krommer, Christoph W. Ueberhuber. SIAM, 1998, xix + 445 pp, $64 (P). [ISBN 0-89871-374-9] Modern survey of methods and underlying mathematical principles. Gives special attention to recent developments (e.g., parallel algorithms) and to the practical application of theoretical results. Discusses both one- and multi-dimensional formulas. AO

Numerical Analysis, P. *Finite Element Analysis of Acoustic Scattering.* Frank Ihlenburg. Appl. Math. Sci., V. 132. Springer-Verlag, 1998, xiv + 224 pp, $59.95. [ISBN 0-387-98319-8]

Numerical Analysis, P. *Combined Methods for Elliptic Equations with Singularities, Interfaces and Infinities.* Zi Cai Li. Math. & Its Applic., V. 444. Kluwer Academic, 1998, xxiv + 476 pp, $214. [ISBN 0-7923-5084-7]

Functional Analysis, P. *Morita Equivalence and Continuous-Trace C*-Algebras.* Iain Raeburn, Dana P. Williams. Math. Surv. & Mono., V. 60. AMS, 1998, xiv + 327 pp, $65. [ISBN 0-8218-0860-5]

Analysis, P. *Weyl Transforms.* M.W. Wong. Universitext. Springer-Verlag, 1998, viii + 158 pp, $44.95. [ISBN 0-387-98414-3]

Analysis, P. *Holomorphic Spaces.* Eds: Sheldon Axler, John E. McCarthy, Donald Sarason. Math. Sci. Res. Inst., V. 33. Cambridge Univ Pr, 1998, ix + 476 pp, $54.95. [ISBN 0-521-63193-9] 16 expository papers presented in 1995 at MSRI. Topics: Bergman spaces; Hankel operators; the Dirichlet space; subnormal operators; operator models, interpolation problems, systems theory.

Analysis, P. *Boundary Value Problems for Transport Equations.* Valeri Agoshkov. Modeling & Simul. in Sci., Eng., & Tech. Birkhäuser

Boston, 1998, xvii + 278 pp, $69.95. [ISBN 0-8176-3986-1]

Analysis, P. *Stability of Functional Equations in Several Variables.* Donald H. Hyers, George Isac, Themistocles M. Rassias. Prog. in Nonlinear Diff. Equat. & Their Applic., V. 34. Birkhäuser Boston, 1998, vii + 313 pp, $79.50. [ISBN 0-8176-4024-X]

Algebraic Geometry, P. *Algebraic Groups and Their Birational Invariants.* V.E. Voskresenskiĭ. Transl. of Math. Mono., V. 179. AMS, 1998, xii + 218 pp, $99. [ISBN 0-8218-0905-9]

Algebraic Geometry, T(16: 2). *Using Algebraic Geometry.* David Cox, John Little, Donal O'Shea. Grad. Texts in Math., V. 185. Springer-Verlag, 1998, xii + 499 pp, $59.95. [ISBN 0-387-98487-9] Overview of recent developments in modern, computational algebraic geometry (using Gröbner bases, resultants, etc.). Illustrates applications in combinatorics, local rings, algebraic coding theory. RM

Differential Geometry, T(15–17: 3), P, L. *Modern Differential Geometry of Curves and Surfaces with Mathematica, Second Edition.* Alfred Gray. CRC Pr, 1998, xxiv + 1053 pp. [ISBN 0-8493-7164-3] New chapters on global curve theory, space curves, minimal surfaces, inversions, cyclides, the Gauss–Bonnet Theorem, and global surface theory. (*First Edition*, Extended Review, December 1995.)

Differential Geometry, P. *Compactifications of Symmetric Spaces.* Yves Guivarc'h, Lizhen Ji, J.C. Taylor. Progress in Math., V. 156. Birkhäuser Boston, 1998, xi + 284 pp, $54.95. [ISBN 0-8176-3899-7]

Geometry, T(13–14: 2), L. *The Geometric Viewpoint: A Survey of Geometries.* Thomas Q. Sibley. Addison-Wesley, 1998, ix + 309 pp, [ISBN 0-201-87450-4]; *Instructor's Resource Guide: Including Solutions and Projects for the Geometer's Sketchpad to Accompany*, iii + 68 pp. [ISBN 0-201-32550-0] Readable introduction to and survey of modern geometry including transformations, symmetry, projective, finite, and non-Euclidean geometries. Examples and applications nicely motivate the development; interesting exercises. RM

Geometry, P, L. *Non-Euclidean Geometry, Sixth Edition.* H.S.M. Coxeter. MAA, 1998, xviii + 336 pp, $30.95 (P). [ISBN 0-88385-522-4] This republication of a classic text includes a new section on the concept of inversive distance.

Algebraic Topology, P. *Homotopy Theory via Algebraic Geometry and Group Representations.* Eds: Mark Mahowald, Stewart Priddy. Contemp. Math., V. 220. AMS, 1998, xi + 379 pp, $74 (P). [ISBN 0-8218-0805-2] Proceedings of a 1997 conference at Northwestern University.

Optimization, P. *Advances in Nonlinear Programming.* Ed: Ya-xiang Yuan. Appl. Optim., V. 14. Kluwer Academic, 1998, xiii + 351 pp, $149. [ISBN 0-7923-5053-7] Proceedings of a 1996 conference in Beijing, China, honoring M.J.D. Powell's 60th birthday. Includes Powell's keynote address, 8 invited papers, and 9 contributed papers.

Optimization, P. *Optimisation et analyse convexe.* Jean-Baptiste Hiriart-Urruty. Presses Universitaires de France, 1998, 376 pp, 198 FF (P). [ISBN 2-13-048983-4]

Optimization, P. *Quasiconvex Optimization and Location Theory.* Jaoquim António dos Santos Gromicho. Appl. Optim., V. 9. Kluwer Academic, 1998, xxi + 218 pp, $109. [ISBN 0-7923-4694-7]

Optimal Control, P. *Lecture Notes in Control and Information Sciences–235: H_∞ Control and Its Applications.* Ben M. Chen. Springer-Verlag, 1998, xi + 351 pp, $92 (P). [ISBN 1-85233-026-0]

Optimal Control, P. *Calculus of Variations and Optimal Control.* A.A. Milyutin, N.P. Osmolovskii. Transl. of Math. Mono., V. 180. AMS, 1998, xii + 372 pp, $129. [ISBN 0-8218-0753-6]

Stochastic Processes, P. *Random Fields and Stochastic Partial Differential Equations.* Yu. A. Rozanov. Math. & Its Applic., V. 438. Kluwer Academic, 1998, vii + 229 pp, $105. [ISBN 0-7923-4984-9]

Elementary Statistics, T(13: 2), C, L. *Modern Engineering Statistics.* Lawrence L. Lapin. Duxbury Pr, 1997, xv + 583 pp, $78.95, with disk. [ISBN 0-534-50883-9] Emphasizes applications and key statistical concepts. Favors insight over rigor; practicality over mathematical elegance; simplicity over formalism. Many engineering applications. Topics: descriptive statistics, data analysis, statistical process control, statistical models, probability, sampling distributions, statistical estimation, statistical testing, regression analysis, analysis of variance, and experimental design. Many exercises, some answers. KB

Elementary Statistics, T(13: 2), C. *Exploring Statistics: A Modern Introduction to Data Analysis and Inference, Second Edition.* Larry J. Kitchens. Duxbury Pr, 1998, xv + 940 pp, $73.95, with disk. [ISBN 0-534-26346-1] Updated data sets, exercises, and examples

(over 1400 exercises and 400 data sets on disk). Emphasizes analysis and interpretation of data rather than calculations. Actively involves student in the learning process. (*First Edition*, TR, March 1988.) KB

Statistical Methods, P. *Additive and Multiplicative Semiparametric Models in Accelerated Life Testing and Survival Analysis.* V. Bagdonavičius, M. Nikulin. Papers in Pure & Appl. Math., V. 108. Queen's Univ, 1998, ii + 109 pp, (P). [ISBN 0-88911-802-7]

Statistical Methods, P. *Statistical Design and Analysis of Experiments.* Peter W.M. John. Classics in Appl. Math., V. 22. SIAM, 1998, xxiv + 356 pp, $40 (P). [ISBN 0-89871-427-3] Unabridged republication of the 1971 Macmillan Publishing Co. edition (TR, May 1971). New Preface describes changes in the field since the book was first published.

Statistical Methods, T(13, 17: 2), P, L. *A First Course in Multivariate Statistics.* Bernard Flury. Texts in Stat. Springer-Verlag, 1997, xv + 713 pp, $79.95. [ISBN 0-387-98206-X] Unified treatment of theoretical and practical aspects of multivariate statistics. Assumes knowledge of basic statistics and matrix algebra. Topics: multivariate normal distribution, parameter estimation, statistical inference for means, classification, multivariate analysis of variance, logistic regression, principal components, and normal mixtures. KB

Mathematical Computing, S, P, L. *The MATLAB 5 Handbook.* Darren Redfern, Colin Campbell. Springer-Verlag, 1998, xi + 488 pp, $34.95 (P). [ISBN 0-387-94200-9] For reference rather than tutorial use. Organized by problem category (e.g., linear equations, ordinary differential equations, animation). Entries give command name, parameters, hints, and cross-references.

Mathematical Computing, S, P. *The Mathematica Primer.* Kevin R. Coombes, *et al.* Cambridge Univ Pr, 1998, xvii + 214 pp, $24.95 (P); $64.95. [ISBN 0-521-63715-5; 0-521-63130-0] A brief introduction to *Mathematica* Version 3 for novice users. AO

Applications (Economics), P, L. *Methods of Mathematical Finance.* Ioannis Karatzas, Steven E. Shreve. Applic. of Math., V. 39. Springer-Verlag, 1998, xv + 407 pp, $69.95. [ISBN 0-387-94839-2] Sets up Brownian model for financial markets; treats pricing and hedging contingent claims; addresses problems faced by agents (one or several) for optimal consumption and investment decisions. Reader must be familiar with probability and stochastic processes. KS

Applications (Fluid Mechanics), P. *Weakly Nonlocal Solitary Waves and Beyond-All-Orders Asymptotics: Generalized Solitons and Hyperasymptotic Perturbation Theory.* John P. Boyd. Math. & Its Applic., V. 442. Kluwer Academic, 1998, xix + 590 pp, $198. [ISBN 0-7923-5072-3]

Applications (Mechanics), P. *Computational Inelasticity.* J.C. Simo, T.J.R. Hughes. Interdisc. Appl. Math., V. 7. Springer-Verlag, 1998, xiv + 392 pp, $59.95. [ISBN 0-387-97520-9]

Applications (Mechanics), P. *Generalized Analytic Functions: Theory and Applications to Mechanics.* Eds: Helmut Florian, *et al.* Intern. Soc. for Analysis, Applic. & Comput., V. 1. Kluwer Academic, 1998, xxvi + 311 pp, $140. [ISBN 0-7923-5043-X] Proceedings of a 1997 conference at the Technical University of Graz. First part treats generalized analytic functions in the complex plane and in higher dimensions. Second part focuses on applications to mechanics.

Applications (Physics), T(17: 1). *The Geometry of Physics: An Introduction.* Theodore Frankel. Cambridge Univ Pr, 1997, xxii + 654 pp, $95. [ISBN 0-521-38334-X] Introduction to differential geometry and topology, Lie groups, bundles and forms, etc., for modern and classical physics. Good, intuitive development of the fundamental concepts by discussion of surfaces in space. RM

Applications (Systems Theory), P. *Lecture Notes in Control and Information Sciences–234: Neural Networks in Multidimensional Domains: Fundamentals and New Trends in Modelling and Control.* Paolo Arena, *et al.* Springer-Verlag, 1998, xv + 165 pp, $59 (P). [ISBN 1-85233-006-6]

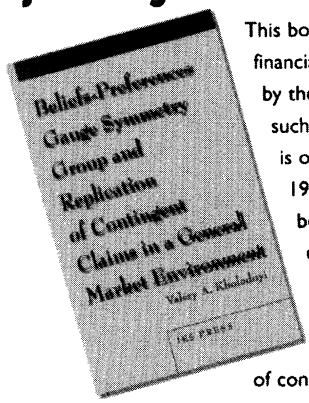
Applications (Systems Theory), P. *Lecture Notes in Control and Information Sciences–232: Experimental Robotics V.* Eds: Alicia Casals, Anibal T. de Almeida. Springer-Verlag, 1998, xix + 700 pp, $129 (P). [ISBN 3-540-76218-3] Proceedings of a 1997 conference in Barcelona, Spain.

Applications, P. *Optical Pattern Recognition.* Eds: Francis T.S. Yu, Suganda Jutamulia. Cambridge Univ Pr, 1998, xv + 440 pp, $105. [ISBN 0-521-46517-6] 15 chapters, each written by a specialist, provide an overview of theoretical and practical aspects of the subject.

Reviewers

KB: Karla Ballman, Macalester; RM: Richard Molnar, Macalester; AO: Arnold Ostebee, St. Olaf; KS: Karen Saxe, Macalester.

Beliefs-Preferences Gauge Symmetry Group and Replication of Contingent Claims in a General Market Environment



This book studies the actual financial phenomena underlying the evaluation of financial derivatives, which is today virtually identified with and even replaced by the study of the mathematical aspects of stochastic calculus as a model for such phenomena. It expresses the view that the study of financial phenomena is on the brink of a revolution similar to that of quantum physics in the 1920s. A fundamental symmetry, a gauge symmetry, is introduced between beliefs of market participants and their preferences in a general market environment for a market with exchange of an arbitrary number of arbitrary underlying securities. The practical applications of this gauge symmetry range from the detection of a new type of true arbitrage to the beliefs-preferences-independent valuation and dynamic replication of contingent claims in a general market environment.

To Order Call: 1-888-437-4979 • Special Introductory Price $149.00

ISBN 0-9663032-1-0, Hardcover, 460 Pages. • Additional Discount Available for Full-time Students

IES Press • P.O. Box 14008 • Research Triangle Park, NC • 27709

For Other IES Press Publications, See Web Site: www.ieslc.com

THE MATHEMATICAL ASSOCIATION OF AMERICA

Julia a life in mathematics

Constance Reid

Constance Reid, an established writer about mathematicians, has written an excellent and loving book, about her sister Julia Robinson, the mathematician. The author has written that she wants the book to be one for all age groups and she has succeeded admirably in making it so...Julia wanted to be known as a mathematician, not a woman mathematician and rightly so! However, she was, and is, a wonderful role model for women aspiring to be mathematician. What a great gift this book would be!

—Alice Schafer, Former President, AWM

This book is a small treasure, one which I want to share with all my mathematical friends. The assembly of several articles and additional photos and remarks provides the image of a mathematician of extraordinary taste, tenacity and generosity.... Julia Robinson broke ground in displaying the deep connections between number theory and logic. Her results have led to a very active area today, making the appearance of this book very timely. Her work and her example are however timeless and I can think of no better advice to give a young mathematician, either in how to do mathematics. or how to behave in mathematics, than: "Be like Julia!"

—Carol Wood, Deputy Director, MSRI

Julia is the story of the life of Julia Bowman Robinson, the gifted and highly original mathematician who during her lifetime was recognized in ways that no other woman mathematician had been recognized up to that time. In 1976 she became the first woman mathematician elected to the National Academy of Sciences and in 1983 the first woman elected president of the American Mathematical Society.

This unusual book, profusely illustrated with previously unpublished personal and mathematical memorabilia, brings together in one volume the prizewinning "Autobiography of Julia Robinson" by her sister, the popular mathematical biographer Constance Reid, and three very personal articles about her work by outstanding mathematical colleagues.

All royalties from sales of this book will go to fund a Julia Robinson Prize in Mathematics at the high school from which she graduated.

Catalog Code: JULIA/JR

136 pp., Hardbound, 1996, ISBN 0-88385-520-8

List: $27.00 MAA Member: $20.00

Phone in Your Order Now! ☎ 1-800-331-1622

Prentice Hall Advanced Mathematics

New Titles

Introduction to Analysis
Arthur Mattuck, MIT

Multivariable Calculus with Vectors
Hartley Rogers, Jr., MIT

Differential Equations: Modeling with MATLAB®
Paul Davis, Worchester Polytechnic Institute
Cleve Moler

Discrete Mathematics, Fourth Edition
Ken Ross and Charles Wright, both of the University of Oregon

Abstract Algebra, Second Edition
David Dummit and Richard Foote, both of the University of Vermont

Numerical Methods Using MATLAB®, Third Edition
John Mathews, California State University at Fullerton
Kurtis Fink, Northwest State University

A Contextual History of Mathematics
Ronald Calinger, Catholic University of America

Introduction to Mathematical Programming
Russell Walker, Carnegie-Mellon University

Differential Equations and Linear Algebra, Second Edition
Stephen Goode, California State University at Fullerton

Vector Calculus, Linear Algebra, and Differential Form: A Unified Approach
John Hubbard, Cornell University
Barbara Hubbard

Advanced Calculus: A Friendly Approach
Witold Kosmala, Appalachian State University

Introductory Combinatorics, Third Edition
Richard Brualdi, University of Wisconsin

Foundations of Plane Geometry
Harvey Blau, Northern Illinois University
John Wetzel, University of Illinois, Urbana

Introduction to Topology
Dennis Roseman, University of Iowa

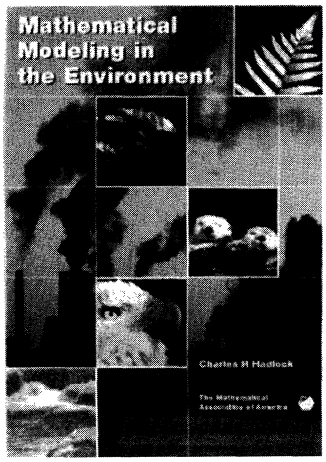
Partial Differential Equations: Sources and Solutions
Arthur David Snider, University of South Florida

Principles of Mathematical Problem Solving
Martin Erickson and Joe Flowers, both of Truman State University

A First Course in Fourier Analysis
David Kammler, Southern Illinois University

Calculus with Early Vectors
by Phillip Zenor, Edward Slaminka, and Donald Thaxton, all of Auburn University
Starting with an introduction to vectors in Chapter 1 and integrating this topic (and differential equations) throughout, this text is written for students taking a concurrent calculus-based physics course. Physics and engineering applications receive extra emphasis.

Visit us on the Web at www.prenhall.com



THE MATHEMATICAL ASSOCIATION OF AMERICA

Mathematical Modeling in the Environment

Charles Hadlock

Series: Classroom Resource Materials

Packaged with a PC compatible disk that enhances the material in the text.

Suitable for classroom adoption in an innovative course for

- **a general education mathematics elective**
- **a mathematics or science major advanced elective**
- **an interdisciplinary course, even at a relatively elementary level**
- **a mathematical modeling course in a civil/environmental engineering program**

This book has a dual objective: first, to introduce the reader to some of the most important and widespread environmental issues of the day; and second, to illustrate the vital role played by mathematical models in investigating these issues. The environmental issues addressed include: ground-water contamination, air pollution, and hazardous material emergencies. These issues are presented in their full real-world context, not as scientific or mathematical abstractions; and for background, readers are invited to investigate their status in their own communities.

The first part of the book leads the reader through relatively elementary modeling of these phenomena, including simple algebraic equations for ground water, slightly more complex algebraic equations (preferably implemented on a spreadsheet or other computerized framework) for air pollution, and a fully computerized modeling package for hazardous materials incident analysis. The interplay between physical intuition and mathematical analysis is emphasized.

For more advanced readers, the second part of the book returns to the same three subjects but with a higher level of mathematical sophistication (adjustable to the preparation of the reader by selection of subsections.) Many important classical mathematical themes are developed through this context, examples coming from single and multivariable calculus, differential equations, numerical analysis, linear algebra and probability. The material is presented in such a way as to minimize the required background and to encourage the subsequent study of some of these fields.

An elementary course for a general audience could be based entirely on Part I, and a higher level mathematics, science, or engineering course could move quickly to Part 2.

A PC compatible diskette packaged with the text contains a spreadsheet program that facilitates the numerical experimentation with the Gaussian plume equation introduced in Chapter 3, as well as public domain DOS program (ARCHIE) for evaluating the consequences from various hazardous materials scenarios (e.g., the physical extent of flammable and toxic vapor clouds). Text is not tied to the use of this software, but it is included as an aid to meet the pedagogical objectives of the text.

Catalog Code: ENV/SA
312 pp., Paperbound, 1998, ISBN 0-88385-709-X
List: $55.00 MAA Member: $43.95

Instructor's and Solutions Manual for *Mathematical Modeling in the Environment*

Charles Hadlock

Contains the complete solutions and further discussion of nearly every exercise presented in the textbook. This includes both the mathematical/computational exercises as well as the research questions and investigations. Readers will benefit greatly from perusing solutions to the problems whether they have worked them out themselves or not. Students using this volume will still need to work out solutions of research questions using their own sources and adapting them to their own geographic locations, or using their own computational schemes, so this volume could well be useful for students in many course contexts. Enrichment material is included on the topics of some of the exercises. Advice for teachers who lack previous environmental experience, but who want to teach this material is also provided and makes it practical for such persons to offer a course based on these volumes.

Catalog Code: EVS/SA
150 pp., Paperbound, 1998, ISBN 0-88385-713-8
List: $18.95 MAA Member: $14.95

Phone in Your Order Now! ☎ *1-800-331-1622*

THE MATHEMATICAL ASSOCIATION OF AMERICA

New from William Dunham, award-winning author of Journey through Genius: The Great Theorems of Mathematics, and The Mathematical Universe....

Euler

The Master of Us All

William Dunham

Series: Dolciani Mathematical Expositions

Without question, Leonhard Euler (1707-1783) ranks among history's greatest mathematicians. Across six decades of unmatched productivity, and despite a visual impairment that grew ever worse, he charted the course of mathematics throughout the eighteenth century and beyond. His reputation is captured in Laplace's famous admonition, "Read Euler, read Euler. He is the master of us all."

Written for the mathematically literate reader, this book provides a glimpse of Euler in action. Following an introductory biographical sketch are chapters describing his contributions to eight different topics—number theory, logarithms, infinite series, analytic number theory, complex variables, algebra, geometry, and combinatorics. Each chapter begins with a prologue to establish the historical context and then proceeds to a detailed consideration of one or more Eulerian theorems on the subject at hand. Each chapter concludes with an epilogue surveying subsequent developments or addressing related questions that remain unanswered to this day. At the end of the book is a brief outline of Euler's collected works, the monumental *Opera Omnia*, whose publication has consumed virtually all of the twentieth century.

In all, the book contains three dozen proofs from this remarkable individual. Yet this is merely the tip of the scholarly iceberg, for Euler produced over 30,000 pages of pure and applied mathematics during his lifetime. *Euler: The Master of Us All* samples the work of a mathematician whose influence, industry, and ingenuity are of the very highest order.

Catalog Code: DOL-22/JR
192 pp., Paperbound, ISBN- 0-88385-328-0
List: $29.95 MAA Member: $23.95

Phone in Your Order Now! ☎ *1-800-331-1622*

Monday – Friday 8:30 am – 5:00 pm FAX (301) 206-9789

or mail to: The Mathematical Association of America, PO Box 91112, Washington, DC 20090-1112

Shipping and Handling: Postage and handling are charged as follows: **USA orders (shipped via UPS):** $2.95 for the first book, and $1.00 for each additional book. **Canadian orders:** $4.50 for the first book and $1.50 for each additional book. Canadian orders will be shipped within 10 days of receipt of order via the fastest available route. We do not ship via UPS into Canada unless the customer specially requests this service. Canadian customers who request UPS shipment will be billed an additional 7% of their total order. **Overseas orders:** $3.50 per item ordered for books sent surface mail. Airmail service is available at a rate of $7.00 per book. Foreign orders must be paid in US dollars through a US bank or through a New York clearinghouse. Credit Card orders are accepted for all customers.

Name _____

Address _____

City _____ State _____ Zip _____

Phone _____

Qty.	Catalog Code	Price	Amount
_____	DOL-22/JR	_____	_____

All orders must be prepaid with the exception of books purchased for resale by bookstores and wholesalers.

Shipping & handling _____

TOTAL _____

Payment ☐ Check ☐ VISA ☐ MasterCard

Credit Card No. _____ Expires ___/___

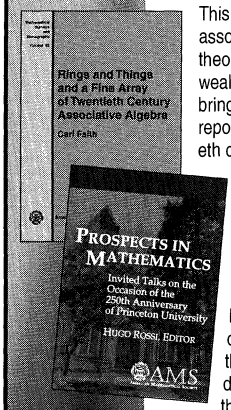
Signature _____

AMERICAN MATHEMATICAL SOCIETY

New Titles from the AMS

Rings and Things and a Fine Array of Twentieth Century Associative Algebra

Carl Faith, *Professor Emeritus, Rutgers University, New Brunswick, NJ*



This book surveys more than 125 years of aspects of associative algebras, especially ring and module theory. It is the first to probe so extensively such a wealth of historical development. Moreover, the author brings the reader up to date, in particular through his report on the subject in the second half of the twentieth century.

Included in the book are certain categorical properties from theorems of Frobenius and Stickelberger on the primary decomposition of finite Abelian groups; Hilbert's basis theorem and his Nullstellensatz, including the modern formulations of the latter by Krull, Goldman, and others; Maschke's theorem on the representation theory of finite groups over a field; and the fundamental theorems of Wedderburn on the structure of finite dimensional algebras and finite skew fields and their extensions by Braver, Kaplansky, Chevalley, Goldie, and others. A special feature of the book is the in-depth study of rings with chain condition on annihilator ideals pioneered by Noether, Artin, and Jacobson and refined and extended by many later mathematicians.

Two of the author's prior works, *Algebra: Rings, Modules and Categories, I* and *II* (Springer-Verlag, 1973), are devoted to the development of modern associative algebra and ring and module theory. Those works serve as a foundation for the present survey, which includes a bibliography of over 1,600 references and is exhaustively indexed.

In addition to the mathematical survey, the author gives candid and descriptive impressions of the last half of the twentieth century in "Part II: Snapshots of Some Mathematical Friends and Places". Beginning with his teachers and fellow graduate students at the University of Kentucky and at Purdue, Faith discusses his Fulbright-Nato Postdoctoral at Heidelberg and at the Institute for Advanced Study (IAS) at Princeton, his year as a visiting scholar at Berkeley, and the many acquaintances he met there and in subsequent travels in India, Europe, and most recently, Barcelona.

Comments on the book:

Researchers in algebra should find it both enjoyable to read and very useful in their work. In all cases, [Faith] cites full references as to the origin and development of the theorem .... I know of no other work in print which does this as thoroughly and as broadly.

—John O'Neill, University of Detroit at Mercy

"Part II: Snapshots of Some Mathematical Friends and Places" is wonderful! [It is] a joy to read! Mathematicians of my age and younger will relish reading "Snapshots".

—James A. Huckaba, University of Missouri-Columbia

Mathematical Surveys and Monographs, Volume 65; 1999; 420 pages; Hardcover; ISBN 0-8218-0993-8; List $99; Individual member $59; Order code SURV/65MM91

Prospects in Mathematics

Invited Talks on the Occasion of the 250th Anniversary of Princeton University

Hugo Rossi, *Mathematical Sciences Research Institute, Berkeley, CA*, Editor

In celebration of Princeton University's 250th anniversary, the mathematics department held a conference entitled "Prospects in Mathematics". The purpose of the conference was to speculate on future directions of research in mathematics.

This collection of articles provides a rich panorama of current mathematical activity in many research areas. From Gromov's lecture on quantitative differential topology to Witten's discussion of string theory, new ideas and techniques transfixed the audience of international mathematicians. The volume contains 11 articles by leading mathematicians, including historical presentations by J. Milnor and D. Spencer. It provides a guide to some of the most significant mathematical work of the past decade.

Cover picture of Old Fine Hall at Princeton University is courtesy of Robert P. Matthews, Communications Department, Princeton University.

1999; 154 pages; Hardcover; ISBN 0-8218-0975-X; List $29; All AMS members $23; Order code PIMMM91

M-Theory

AMS VIDEO

Edward Witten, *Institute for Advanced Study, Princeton, NJ*

The problem of unifying quantum mechanics and gravity in a single coherent theory represents an enormous obstacle to full understanding of the forces of nature. The mysterious M-theory has emerged as a likely candidate for such a unifying theory. Whether the "M" stands for marvel or matrix, magic or membrane, it is clear that this area of research is among the most exciting and most profound in all of science today. Edward Witten, one of the world's boldest innovators in this field, provides insights into these extraordinary developments in a completely expository presentation. Students and researchers specializing in mathematics and physics will find this lecture especially appealing. However, because it is completely nontechnical, large parts of it can easily be appreciated by viewers with little or no scientific or mathematical training.

1998; NTSC format on one-half inch VHS videotape, approximately 60 minutes; ISBN 0-8218-1350-1; List $54.95; Individual member $34.95; Order code VIDEO/101MM91

AMS
AMERICAN MATHEMATICAL SOCIETY

All prices subject to change. Charges for delivery are $3.00 per order. For optional air delivery outside of the continental U. S., please include $6.50 per item. *Prepayment required.* Order from: **American Mathematical Society**, P. O. Box 5904, Boston, MA 02206-5904, USA. For credit card orders, fax 1-401-455-4046 or call toll free 1-800-321-4AMS (4267) in the U. S. and Canada, 1-401-455-4000 worldwide. Or place your order through the AMS bookstore at www.ams.org/bookstore. Residents of Canada, please include 7% GST.

SPRINGER FOR MATHEMATICS

PAULO RIBENBOIM, Queen's University, Kingston, Ontario, Canada

FERMAT'S LAST THEOREM FOR AMATEURS

In 1995, Andrew Wiles completed a proof of Fermat's last theorem. Although this was certainly a great mathematical feat, one shouldn't dismiss earlier attempts made by clever amateurs and famous mathematicians alike to solve the problems. In this book, aimed at amateurs curious about the unfolding of the subject, the author restricts his attention exclusively to elementary methods which may only have led to partial solutions, but their interest goes beyond Fermat's problem.

1999/APP. 376 PP./HARDCOVER/$39.95/ISBN 0-387-98508-5

E. KAMERICH, Catholic University of Nijmegen, The Netherlands

A GUIDE TO MAPLE®

This "hands-on" book is for people who are interested in immediately putting Maple to work. The reader is provided with a compact, fast, and surveyable guide that introduces them to the extensive capabilities of the software. The book is sufficient for standard use of Maple and will provide techniques for extending Maple for more specialized work. The author discusses the reliability of results systematically and presents ways of testing questionable results. The book allows a reader to become a user almost immediately and helps him/her to grow gradually to a broader and more proficient use. As a consequence, some subjects are dealt with in an introductory way early in the book, with references to a more detailed discussion later on.

1998/352 PP., 41 ILLUS./HARDCOVER/$39.95
ISBN 0-387-94116-9

JOE MAZUR, Marlboro College, VT

EXPLORATIONS IN CALCULUS

Explorations in Calculus is a completely self-contained, cross platform CD tutorial package and electronic study guide for students taking calculus at the college or high school level. Among the many features of the CD are: sounds and animations, text files, examples and exercises, a drawing program, progress checks, feedback, hints to problem solving, and cut and paste and notepad capabilities.

1999/APP. 24 PP., 68 ILLUS./$37.95 (TENT.)
ISBN 0-387-14249-5
TEXTS IN MATHEMATICAL SCIENCES

Now available bundled with Student Minitab!

ALLAN ROSSMAN, Dickinson College, Carlisle, PA and **BETH L. CHANCE,** University of the Pacific, Stockton, CA

WORKSHOP STATISTICS

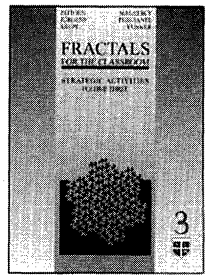
Student Minitab Version

1999/152 PP./SOFTCOVER/$59.95 (TENT.)
ISBN 0-387-91580-X

HEINZ-OTTO PEITGEN and **HARTMUT JÜRGENS,** both of University of Bremen, Germany, **DIETMAR SAUPE,** University of Freiburg, Germany, **EVAN MALETSKY,** Montclair State University, NJ and **TERRY PERCIANTE,** Wheaton College, IL

FRACTALS FOR THE CLASSROOM STRATEGIC ACTIVITIES

Volume 3



This third and final volume of *Strategic Activities* focuses on fractal images and the mechanisms by which they are formed. The general pattern and specific steps used to construct a fractal image illustrated throughout this volume comprise an iterated function system. The objective of this volume is to investigate the processes and often surprising results of applying such systems. The contents of this volume joined with the details contained in the prior two books provide a comprehensive survey of fractal geometry and chaos theory. In addition to the notions of this new and emerging discipline, the dynamic nature of the research and the experimental characteristics of related applications provide an engaging paradigm for classroom activity.

Contents: Connections to the Curriculum • Foreword (by Jana Wallace) • IFS in Two Dimensions • IFS and Geometric Genetic Codes • Answers

1998/APP. 121 PP., 285 ILLUS./SOFTCOVER/$24.95 (TENT.)
ISBN 0-387-98420-8

Order Today!

Call: 1-800-SPRINGER or **Fax:** (201)-348-4505
Write: Springer-Verlag New York, Inc., Dept. S273, PO Box 2485, Secaucus, NJ 07096-2485
Visit: Your local technical bookstore
E-mail: orders@springer-ny.com
Instructors: Call or write for info on textbook exam copies
YOUR 30-DAY RETURN PRIVILEGE IS ALWAYS GUARANTEED!

Springer

http://www.springer-ny.com

1/99 Promotion #S273

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, DC 20036